

# Computer Simulation of Molecular Dynamics: Methodology, Applications, and Perspectives in Chemistry

By Wilfred F. van Gunsteren\* and Herman J. C. Berendsen\*

During recent decades it has become feasible to simulate the dynamics of molecular systems on a computer. The method of molecular dynamics (MD) solves *Newton's* equations of motion for a molecular system, which results in trajectories for all atoms in the system. From these atomic trajectories a variety of properties can be calculated. The aim of computer simulations of molecular systems is to compute *macroscopic* behavior from *microscopic* interactions. The main contributions a microscopic consideration can offer are (1) the understanding and (2) interpretation of experimental results, (3) semiquantitative estimates of experimental results, and (4) the capability to interpolate or extrapolate experimental data into regions that are only difficultly accessible in the laboratory. One of the two basic problems in the field of molecular modeling and simulation is how to efficiently search the vast configuration space which is spanned by all possible molecular conformations for the global low (free) energy regions which will be populated by a molecular system in thermal equilibrium. The other basic problem is the derivation of a sufficiently accurate interaction energy function or force field for the molecular system of interest. An important part of the art of computer simulation is to choose the unavoidable assumptions, approximations and simplifications of the molecular model and computational procedure such that their contributions to the overall inaccuracy are of comparable size, without affecting significantly the property of interest. Methodology and some practical applications of computer simulation in the field of (bio)chemistry will be reviewed.

## 1. Introduction

Computational chemistry is a branch of chemistry that enjoys a growing interest from experimental chemists. In this discipline chemical problems are resolved by computational methods. A model of the real world is constructed, both measurable and unmeasurable properties are computed, and the former are compared with experimentally determined properties. This comparison validates or invalidates the model that is used. In the former case the model may be used to study relationships between model parameters and assumptions or to predict unknown or unmeasurable quantities.

Since chemistry concerns the study of properties of substances or molecular systems in terms of atoms, the basic challenge facing computational chemistry is to describe or even predict

1. the structure and stability of a molecular system,
2. the (free) energy of different states of a molecular system,
3. reaction processes within molecular systems

in terms of interactions at the atomic level. These three basic challenges are listed according to increasing difficulty. The first challenge concerns prediction of which state of a system has the lowest energy. The second challenge goes further; it involves prediction of the relative (free) energy of different states. The third challenge involves prediction of the dynamic process of change of states.

Chemical systems are generally too inhomogeneous and complex to be treated by analytical theoretical methods. This is illustrated in Figure 1. The treatment of molecular systems in the gas phase by quantum mechanical methods is straightforward; if a classical statistical mechanical approximation is permitted the problem becomes even trivial. This is due to the possibility of reducing the many-particle problem to a few-particle one based on the low density of a system in the gas phase. In the crystalline solid state, treatment by quantum mechanical or classical mechanical methods is made possible by a reduction of the many-particle problem to a few-(quasi)particle problem based on symmetry properties of the solid state. Between these two extremes, that is, for liquids, macromolecules, solutions, amorphous solids, etc., one is faced with an *essentially* many-particle system. No simple reduction to a few degrees of freedom is possible, and a full treatment of many degrees of freedom is required in order to adequately describe the properties of molecular systems in the fluid-like state. This state of affairs has two direct consequences when treating fluid-like systems.

1. One has to resort to numerical *simulation* of the behavior of the molecular system on a computer, which
2. produces a *statistical ensemble* of configurations representing the state of the system.

If one is only interested in static equilibrium properties, it suffices to generate an ensemble of equilibrium states, which

[\*] Prof. Dr. W. F. van Gunsteren<sup>1,3</sup>  
Department of Physical Chemistry  
University of Groningen  
Nijenborgh 16, NL-9747 AG Groningen (The Netherlands)  
Department of Physics and Astronomy  
Free University  
De Boelelaan 1081, NL-1081 HV Amsterdam (The Netherlands)  
Prof. Dr. H. J. C. Berendsen  
Department of Physical Chemistry  
University of Groningen  
Nijenborgh 16, NL-9747 AG Groningen (The Netherlands)

[\*] Present address:  
Eidgenössische Technische Hochschule  
ETH Zentrum  
Universitätstrasse 6, CH-8092 (Switzerland)

may lack any temporal correlations. To obtain dynamic and non-equilibrium properties dynamic simulation methods that produce trajectories in phase space are to be used. The connection between the microscopic behavior and macroscopic properties of the molecular system is governed by the laws of *statistical mechanics*.

Figure 1 also shows the broad applicability of computer simulation methods in chemistry. For any fluid-like, essentially many-particle system, it is the method of choice.

|                                       | CRYSTALLINE<br>SOLID STATE | LIQUID STATE<br>MACROMOLECULES | GAS PHASE |
|---------------------------------------|----------------------------|--------------------------------|-----------|
| QUANTUM<br>MECHANICS                  | possible                   | still<br>impossible            | possible  |
| CLASSICAL<br>STATISTICAL<br>MECHANICS | easy                       | computer<br>simulation         | trivial   |

REDUCTION to few degrees of freedom by SYMMETRY ← essential many-particle system → REDUCTION to few particles by DILUTION

Fig. 1. Classification of molecular systems. Systems in the shaded area are amenable to treatment by computer simulation.

The expanding role of computational methods in chemistry has been fueled by the steady and rapid increase in computing power over the last 40 years, as is illustrated in Figure 2. The ratio of performance to price has increased an order of magnitude every 5–7 years, and there is no sign of any weakening in this trend. The introduction of massive parallelism in computer architecture will easily maintain the present growth rate. This means that more complex molecular systems may be simulated over longer periods of time, or that it will be possible to handle more complex interaction functions in the decades to come.

The present article is concerned with computer simulation of molecular systems. In Section 2 the two basic problems are formulated, a brief history of dynamic computer simulation is presented, the reliability of current simulations is discussed, and the usefulness of simulation studies is considered. Section 3 deals with simulation methodology: choice of computational model and atomic interaction function (Section 3.1), techniques to search configuration space for low energy configurations (Section 3.2), boundary conditions (Section 3.3), types of dynamical simulation methods (Sec-

tion 3.4), algorithms for integration of the equations of motion (Section 3.5), and finally equilibration and analysis of molecular systems (Section 3.6). In Section 4, a number of applications of computer simulation in chemistry are dis-

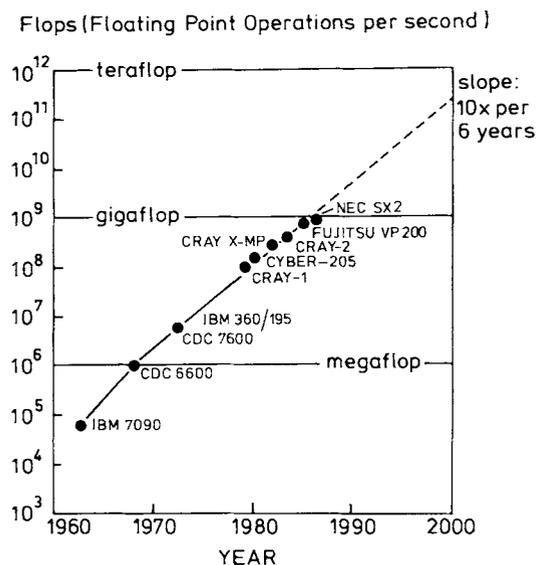


Fig. 2. Development of computing power of the most powerful computers.

cussed and examples are given. Finally, future developments are considered in Section 5. For other relatively recent monographs on computer simulation the reader is referred to the literature (see Refs. [1–9]).

## 2. Computer Simulation of Molecular Systems

### 2.1. Two Basic Problems

Two basic problems are encountered in the computer simulation of fluid-like molecular systems:

1. the *size of the configurational space* that is accessible to the molecular system, and
2. the *accuracy* of the molecular model or atomic interaction function or force field that is used to model the molecular system.



*Wilfred F. van Gunsteren was born in 1947 in Wassenaar (The Netherlands). In 1968 he gained a B.Sc. in physics at the Free University of Amsterdam; in 1976 he was awarded a "Meester" in Law, and in 1976 a Ph.D. in nuclear physics. After postdoc years at the University of Groningen (1976–1978) and at Harvard University (1978–1980) he was, from 1980 until 1987, senior lecturer and, until August 1990, Professor for Physical Chemistry at the University of Groningen. Since September 1987 he has also been Professor of Computer Physics at the Free University Amsterdam. In 1990 he was offered and accepted a professorship of Computer Chemistry at the ETH Zürich. He is holder of a gold medal for research of the Royal Netherlands Chemical Society; in 1988 he was Degussa Guest Professor at the University of Frankfurt. His main interests center on the physical fundamentals of the structure and function of biomolecules.*

Table 1. Models at different levels of approximation.

| Model                         | Degrees of freedom        |                  | Example of                           |                          |   |
|-------------------------------|---------------------------|------------------|--------------------------------------|--------------------------|---|
|                               | left                      | removed          | predictable property                 | force field              |   |
| - quantum mechanical          | nuclei, electrons         | nucleons         | reactions                            | Coulomb                  | increasing:<br>simplicity,<br>speed of computation,<br>search power,<br>time scale<br><br>decreasing:<br>complexity, accuracy of<br>atomic properties |
| - all atoms, polarizable      | atoms<br>dipoles          | electrons        | binding charged<br>ligand            | ionic models [10]        |   |
| - all atoms                   | solute + solvent<br>atoms | dipoles          | hydration                            | OPLS [11]<br>GROMOS [12] |   |
| - all solute atoms            | solute atoms              | solvent          | gas phase conformation               | MM2 [13]                 |   |
| - groups of atoms<br>as balls | atom groups               | individual atoms | folding topology<br>of macromolecule | LW [14]                  |   |

### 2.1.1. Size of the Configurational Space

The simulation of molecular systems at non-zero temperatures requires the generation of a statistically representative set of configurations, a so-called *ensemble*. The properties of a system are defined as ensemble averages or integrals over the configuration space (or more generally phase space). For a many-particle system the averaging or integration will involve many degrees of freedom, and as a result can only be carried out over part of the configuration space. The smaller the configuration space, the better the ensemble average or integral can be approximated. When choosing a model from which a specific property is to be computed, one would like to explicitly include only those degrees of freedom on which the required property is dependent.

In Table 1 a hierarchy of models is shown, in which specific types of degrees of freedom are successively removed. Examples are given of properties that can be computed at the different levels of approximation. If one is interested in chemical reactions, a quantum mechanical treatment of electronic degrees of freedom with the nuclear coordinates as parameters is mandatory. The intranuclear degrees of freedom (nucleon motion) are left out of the model. The computing power required for a quantum mechanical treatment scales at least with the third power of the number of electrons  $N_e$  that are considered. Such a treatment is only possible for a limited number of degrees of freedom. At a next level of approximation the electronic degrees of freedom are removed from the model and approximated by (point) polarizabilities. Such a model, allowing for motions of polarizable atoms, could, e.g., be used to compute the binding properties of polar or charged ligands which will polarize the receptor. At this level of modeling the computing effort scales with the square of the number of atoms  $N_a$ . This means that many more atoms can be considered than at the quantum level. Removal of the polarizability of the atoms yields a next level of approximation, in which only atomic positional degrees of freedom are considered and the mean polarization is included in the effective interatomic interaction function. Such a model allows for the study of solvation properties of molecules. When studying solutions, a next level of approximation is reached by omitting the solvent degrees of freedom from the model and simultaneously adapting the interaction function for the solute such that it includes the mean solvent effect. When studying protein folding, the configuration space spanned by all

protein atoms is still far too large to be searched for low energy conformations. In this case a further reduction in it can be obtained by representing whole groups of atoms, for example an amino acid residue, as one or two balls (3 or 6 degrees of freedom).

From this discussion it is clear that the level of approximation of the model that will be used in the simulation will depend on the specific property one is interested in. The various force fields that are available correspond to different levels of approximation, as is illustrated in the right half of Table 1. There is a hierarchy of force fields.

Once the level of approximation has been chosen, that is, the types of degrees of freedom in the model, one must decide how many degrees of freedom (electrons, atoms, etc.) are to be taken into account. How small a system can be chosen without seriously affecting a proper representation of the property of interest? The smaller the size of the system the better its degrees of freedom can be sampled, or in dynamic terms, the longer the time scale over which it can be simulated.

We may summarize the first basic problem in the computer simulation of molecular systems as follows: the level of approximation of the model should be chosen such that those degrees of freedom that are essential to a proper evaluation of the quantity or property of interest can be sufficiently sampled. In practice any choice involves a compromise between type and number of degrees of freedom and extent of the simulation on the one hand, and the available computing power on the other.

### 2.1.2. Accuracy of Molecular Model and Force Field

When the degrees of freedom are infinitely dense or long sampled the accuracy by which various quantities are predicted by a simulation will depend solely on the quality of the assumptions and approximations of the molecular model and interatomic force field. At the level (Table 1) of quantum mechanical modeling the basic assumption is the validity of the Born-Oppenheimer approximation separating electronic and nuclear motion. The interaction between point atoms and electrons is described by Coulomb's law and the Pauli exclusion principle. When excluding chemical reactions, low temperatures or details of hydrogen atom motion, it is relatively safe to assume that the system is governed by the laws of classical mechanics. The atomic interaction function is called an *effective interaction* since the average effect of the

omitted (electronic) degrees of freedom has been incorporated in the interaction between the (atomic) degrees of freedom explicitly present in the model. To each level of approximation in Table 1 there corresponds a type of effective interaction or force field. For example, a pair potential with an enhanced dipole moment (enhanced atomic charges) may be used as an effective potential that mimics the average effect of polarizability.

In view of the different levels of approximation of molecular models it is not surprising that the literature contains a great variety of force fields. They can be classified along different lines:

- type of compound that is to be mimicked, e.g. carbohydrates, sugars, polypeptides, polynucleotides;
- type of environment of the compound of interest; e.g. gas phase, aqueous or nonpolar solution;
- range of temperatures covered by the effective interaction;
- type of interaction terms in the force field; e.g. bond stretching, bond angle bending, torsional terms, two-, three- or many-body nonbonded terms;
- functional form of the interaction terms; e.g. exponential or 12th power repulsive nonbonded interaction;
- type of parameter fitting, that is, to which quantities are parameters fitted, and are experimentally or ab-initio theoretically obtained values used as target values.

We note that the choice of a particular force field should depend on the system properties one is interested in. Some applications require more refined force fields than others. Moreover, there should be a balance between the level of accuracy or refinement of different parts of a molecular model. Otherwise the computing effort put into a very detailed and accurate part of the calculation may easily be wasted due to the distorting effect of the cruder parts of the model.

### 2.1.3. Assumptions, Approximations and Limitations

When using a particular model to predict the properties of a molecular system, one should be aware of the assumptions, simplifications, approximations and limitations that are implicit in the model. Below, we list the four most important approximations and limitations of classical computer simulation techniques which should be kept in mind when using them.

#### 2.1.3.1. Classical Mechanics of Point Masses

A molecular system is described as a system of point masses moving in an effective potential field, which is generally a conservative field, i.e. it only depends on the instantaneous coordinates of the point masses. The motion of the point masses is with a sufficient degree of accuracy governed by the laws of classical mechanics. These assumptions imply the following restrictions to modeling:

- low-temperature (0–10 K) molecular motion is not adequately described;
- detailed motion of light atoms such as hydrogen atoms is not correctly described even at room temperature;
- description of chemical reactions lies outside the scope of classical simulation methods.

For a short discussion of the inclusion of quantum corrections to a classical treatment or of the quantum (dynamical) simulation techniques that are presently under development, we refer to Section 5.1.

#### 2.1.3.2. System Size or Number of Degrees of Freedom to be Included

Only a rather limited number of atoms can be simulated on a computer. Simulations of liquids typically involve  $10^2$ – $10^3$  atoms, simulations of solutions or crystals of macromolecules about  $10^3$ – $2 \times 10^4$  atoms. Generally, the system size is kept as small as possible in order to allow for a sufficient sampling of the degrees of freedom that are simulated. This means that those degrees of freedom that are not essential to the property one is interested in should be removed from the system. The dependence of the property of interest on the size of the system may give a clue to the minimum number of degrees of freedom required for an adequate simulation of it. The larger the spatial correlation length of the property of interest, the more atoms are to be included in the simulation.

#### 2.1.3.3. Sufficiency of Sampling of Configuration Space or Time Scale of Processes

Computer simulation generates an ensemble of configurations of the system. Whether the generated set of configurations is representative for the state of the system depends on the extent to which the important (generally low energy) parts of the configuration space (or, more generally formulated, phase space) have been sampled. This depends in turn on the sampling algorithm that is applied. This should be able to overcome the multitude of barriers of the multidimensional energy surface of the system. In dynamic simulations, the time scale of the process that is mimicked is limited. Presently, molecular simulations cover time periods of 100–1000 picoseconds. In the case of activated processes longer time scales can be reached by using special tricks.<sup>[15,16]</sup> Essentially slow processes, like the folding of a protein, are still well out of reach of computer simulation. We note that the observation that a property is independent of the length of a simulation is a necessary but *not* a sufficient condition for adequate sampling. The system may just reside for a period longer than the simulation time in a certain area without being effected by a nearby region of much lower energy from which it is separated by a large energy barrier.

#### 2.1.3.4. Accuracy of the Molecular Model and Force Field

As is illustrated in Table 1, there exists a variety of molecular models and force fields, differing in the accuracy by which different physical quantities are modeled. The choice of a particular force field will depend on the property and level of accuracy one is interested in.

When studying a molecular system by computer simulation three factors should be considered (cf. Fig. 3).

1. The properties of the molecular system one is interested in should be listed and the configuration space (or time scale) to be searched for relevant configurations should be estimated.

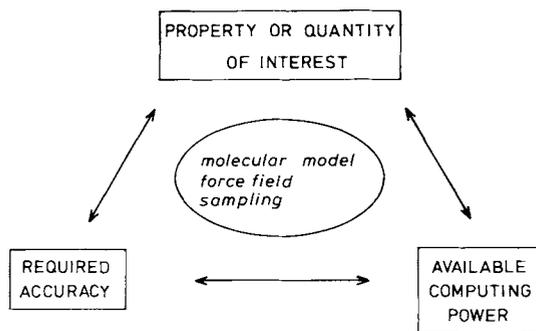


Fig. 3. Choice of molecular model, force field and sample size depends on 1) the property one is interested in (space to be searched), 2) required accuracy of the prediction, 3) the available computing power to generate the ensemble.

2. The required accuracy of the properties should be specified.
3. The available computer time should be estimated.

Given these three specifications the molecular model and force field should be chosen. As is indicated in Table 1 there is a trade off between accuracy of the force field on the one hand and the searching power or time scale that can be attained on the other. We note that in many practical cases one should abandon the idea of trying to simulate the system of interest, since the available computing power will not allow for a sufficiently accurate simulation.

## 2.2. History of Dynamic Computer Simulation

The growth in the number of applications of computer simulation methods in chemistry and physics is directly caused by the rapid increase in computing power over the last four decades (Fig. 2). Table 2 shows the development of

Table 2. History of computer simulation of molecular dynamics.

| Year | System                               | Length of simulation [s] | Required cpu time on super-computer [h] |
|------|--------------------------------------|--------------------------|---|
| 1957 | hard two-dimensional disks [17]      |                          |   |
| 1964 | monatomic liquid [18]                | $10^{-11}$               | 0.05                                    |
| 1971 | molecular liquid [19]                | $5 \times 10^{-12}$      | 1                                       |
| 1971 | molten salt [20]                     | $10^{-11}$               | 1                                       |
| 1975 | simple small polymer [21]            | $10^{-11}$               | 1                                       |
| 1977 | protein in vacuo [22]                | $2 \times 10^{-11}$      | 4                                       |
| 1982 | simple membrane [23]                 | $2 \times 10^{-10}$      | 4                                       |
| 1983 | protein in aqueous crystal [24]      | $2 \times 10^{-11}$      | 30                                      |
| 1986 | DNA in aqueous solution [25]         | $10^{-10}$               | 60                                      |
| 1989 | protein-DNA complex in solution [26] | $10^{-10}$               | 300                                     |
|      | large polymers                       | $10^{-8}$                | $10^3$                                  |
|      | reactions                            | $10^{-4}$                | $10^7$                                  |
|      | macromolecular interactions          | $10^{-3}$                | $10^8$                                  |
|      | protein folding                      | $10^{-1}$                | $10^9$                                  |

the application of molecular dynamics simulation in chemistry. Alder and Wainwright pioneered the method using 2-dimensional hard disks. Rahman, who can really be considered the father of the field, simulated liquid argon in 1964, liquid water in 1971, and superionic conductors in 1978.<sup>[27]</sup> He also made many contributions to the methodology, and stimulat-

ed the dissemination of the technique in the scientific community. In the seventies, the transition from atomic to molecular liquids was made: rigid molecules like water (1971), flexible alkanes (1975), and a small protein, trypsin inhibitor (1977). The application to molten salts (1971) required the development of methods to handle long-range Coulomb interactions. The eighties brought simulations of biomolecules of increasing size in aqueous solution. For a more thorough review of the past and present of dynamic computer simulation in chemistry and physics we refer to the monographs of McCammon and Harvey<sup>[5]</sup> and Allen and Tildesley.<sup>[6]</sup>

From Table 2 it is also clear that a supercomputer is still many orders of magnitude slower than nature: a state of the art simulation is about  $10^{15}$  times slower than nature. Many interesting systems and processes still fall far outside the reach of computer simulation. Yet, with a growth rate of a factor 10 per 5–7 years for (super)computing power the speed of simulation will have caught up with that of nature in about 100 years.

## 2.3. Accuracy and Reliability of Computer Simulation

The reliability of predictions made on the basis of computer simulation will basically depend on two factors: 1) whether the molecular model and force field are sufficiently accurate, and 2) whether the configuration space accessible to the molecular system has been sufficiently thoroughly searched for low-energy configurations. Although the accuracy of a prediction may be estimated by considering the approximations and simplifications of the model and computational procedure, the final test lies in a comparison of theoretically predicted and experimentally measured properties. In order to provide a firm basis for the application of computer simulation methods the results should be compared with experimental data whenever possible.

In this context we would like to stress that good agreement between calculated and experimental data does *not* necessarily mean that the theoretical model underlying the calculation is correct. Good agreement may be due to any of the following reasons:

1. The model is correct, that is, any other assumption used to derive the model, or any other choice of parameter values would give bad agreement with experiment.
2. The property that is compared is insensitive to the assumptions or parameter values of the model, that is, whatever parameter values are used in the model calculation, the agreement with experiment will be good.
3. Compensation of errors occurs, either by chance or by fitting of the model parameters to the desired properties.

A number of examples of the last case, viz. good agreement for the wrong reason, are given in Ref. [28]. Here, we would only like to remind the reader that it is relatively easy, when modeling high-dimensional systems with many parameters, to choose or fit parameters such that good agreement is obtained for a limited number of observable quantities. On the other hand, disagreement between simulation and experiment may also have different causes: 1) the simulation is *not* correct, the experiment is, 2) the simulation is correct, the experiment is *not*.

With these cautionary remarks in mind we would like to turn to an evaluation of the accuracy of current simulation studies by comparing simulated with experimental quantities. Table 3 contains a scheme of the atomic quantities of molecular systems for which comparison between simulated and experimental values is feasible. Four phases are distinguished. The most interesting phase from a chemical point of view is that of a molecule in solution. For this phase only few accurately measured *atomic* properties are available for comparison, leaving thermodynamic system properties like the free energy of solvation as a test ground for the simulation. Considerably more atomic data are available from crystallographic diffraction experiments: atomic positions and mobility, although the accuracy of the latter is much lower than that of the former, due to the simplifying approximations (harmonic isotropic motion) used in the crystallographic refinement process.

Table 3. Possible comparison of simulated properties with experimental ones for complex molecules.

| Atomic properties               | Experimental method                        | Phase                                      |          |          |
|---------------------------------|--|--|----------|----------|
|                                 |  | gas  | solution | membrane |
| <b>Structure</b>                |  |  |          |          |
| - positions                     | { X-ray diffraction<br>neutron diffraction |  |          | ×        |
| - distance                      |  | NMR  | ×        |          |
| - orientations                  | NMR  |  | ×        |          |
| <b>Mobility</b>                 |  |  |          |          |
| - B-factors                     | { X-ray diffraction<br>neutron diffraction |  |          | ×        |
| - occupancy factors             |  | { X-ray diffraction<br>neutron diffraction |          |          |
| <b>Dynamic properties</b>       |  |  |          |          |
| - vibrational frequencies       | infrared spectroscopy                      | ×  |          |          |
| - relaxation rates              | various NMR optical techniques             |  | ×        |          |
| - diffusion                     | NO spin label                              |  | ×        | ×        |
| <b>Thermodynamic properties</b> |  |  |          |          |
| - density                       |  |  | ×        | ×        |
| - free energy                   |  |  | ×        |          |
| - viscosity, conductance        |  |  | ×        |          |

In the following subsections examples will be given of a comparison of various atomic and system properties for different compounds. The examples are taken from our own work, and so are all based on the GROMOS force field.<sup>[1,2]</sup> They only serve as an indication of the degree of accuracy that can be obtained for different properties using current modeling methods.

### 2.3.1. Atomic Properties: Positions and Mobilities in Crystals

From a simulation the average molecular structure can be easily calculated and compared with experiment. Table 4 contains the deviation between simulated and measured atomic positions for a number of molecular crystals. The structural deviation depends on the size of the molecule and ranges from 0.2 Å to 1.2 Å. The numbers are an average over part or over all atoms in the molecules. This means that parts of the molecule will deviate more and other parts will deviate

Table 4. Comparison of MD time-averaged structures with experimental X-ray or neutron diffraction structures.

| Molecule                | Size (number of glucose units or amino acid residues) | Root mean square (RMS) difference in atomic positions [Å] |                         |
|-------------------------|---|---|-------------------------|
|                         |   | C <sub>α</sub> or C <sub>β</sub> atoms                    | all atoms excl. H atoms |
| cyclodextrin (α/β) [29] | 6   | 0.13/0.35   | 0.25/0.51               |
| cyclosporin A [30]      | 11  | 0.3   | 0.6                     |
| trypsin inhibitor [31]  | 58  | 1.0   | 1.5                     |
| subtilisin [32]         | 275   | 1.0   | 1.2                     |

less from the experimental structure. This is illustrated in Figure 4, which shows the deviation for the backbone C<sub>α</sub> atoms as a function of residue number averaged over all four BPTI (bovine pancreatic trypsin inhibitor) molecules in the crystal unit cell. Most of the atoms deviate less than 1 Å. Figure 5 shows the root mean square (rms) atomic positional

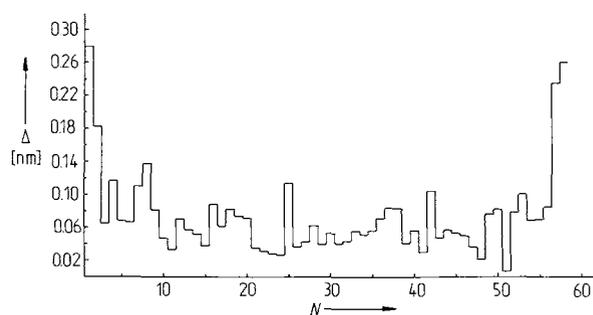


Fig. 4. Root mean square difference between the trypsin inhibitor simulated time- and molecule-averaged C<sub>α</sub> atomic positions and the X-ray positions according to [31]. Δ = root mean square deviation, N = amino acid residue number.

fluctuations calculated from the MD simulation and from the crystallographic B-factors. The largest discrepancy is still rather small, about 0.3 Å.

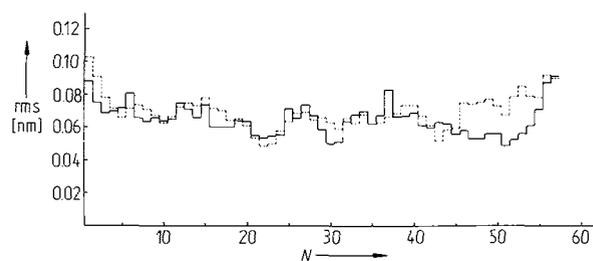


Fig. 5. Root mean square positional fluctuations for the trypsin inhibitor atoms averaged over the backbone atoms of each residue according to [31]. Solid line: simulated results. Broken line: fluctuations derived from a set of X-ray temperature factors. rms = root mean square fluctuation, N = amino acid residue number.

A more stringent test of the simulation is to check whether partially occupied atomic sites are reproduced. The neutron diffraction work on β-cyclodextrin shows that 16 hydrogen atoms occupy two alternative sites. In the crystal simulation a non-zero occupancy is observed for 84% of the hydrogen sites, and for 62% of the hydrogen atoms the relative occu-

pancy of the two alternative sites is qualitatively reproduced.<sup>[29]</sup>

The molecular structure can also be described in terms of hydrogen bonds. Generally the experimentally observed hydrogen bonds are also observed in the simulation. Again cyclodextrin crystals form a sensitive test case for the simulations, since peculiar geometries such as three-center hydrogen bonds, and dynamic processes such as flip-flop hydrogen bonds have been experimentally observed. Almost all experimentally observed three-center hydrogen bonds in crystals of  $\alpha$ - and  $\beta$ -cyclodextrin are reproduced in MD simulation, even as far as the detailed asymmetric geometry is concerned.<sup>[33]</sup> In a so-called flip-flop hydrogen bond the directionality is inverted dynamically. In a MD simulation of  $\beta$ -cyclodextrin, 16 of the 18 experimentally detected flip-flop bonds are reproduced.<sup>[34]</sup>

### 2.3.2. Atomic Properties: Distances in Solution

Structural data of solutions can be obtained by two-dimensional nuclear magnetic resonance spectroscopy (2D-NMR) with exploitation of the nuclear Overhauser (NOE) experiments.<sup>[35]</sup> The data come in the form of a set of upper bounds or constraints to specific proton-proton distances. This affords the possibility of comparing these proton-proton distances as predicted in a simulation with the experimentally measured NOE bounds. In the literature,<sup>[25]</sup> such a comparison is given for a set of 174 NOE's in an eight base-pair DNA fragment in aqueous solution: 80% of the NOE distances are satisfied by the simulation within experimental error, the mean deviation is 0.22 Å, and the maximum deviation amounts to 2.9 Å.

### 2.3.3. Atomic Properties: Orientation of Molecular Fragments in Membranes

The degree of order in a membrane or lipid bilayer can be measured selectively along the aliphatic chain by deuterium NMR spectroscopy.<sup>[36]</sup> In Figure 6<sup>[37]</sup> both experimental and MD order parameters are displayed for all CH<sub>2</sub> units in a sodium ion/water/decanoate/decanol bilayer system as a function of the position of the carbon atom in the aliphatic chain. The experimentally observed plateau in the order parameters is well reproduced.

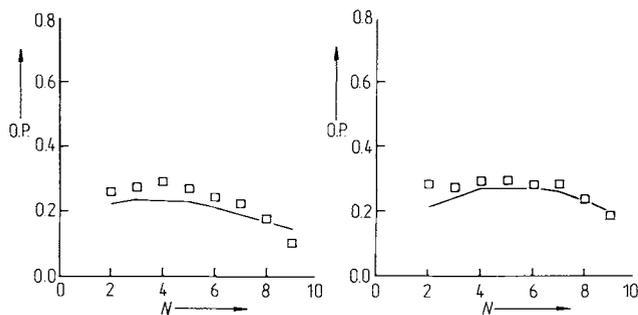


Fig. 6. Carbon-deuterium order parameters as a function of carbon atom number along the aliphatic chain of decanol (upper panel) and decanoate (lower panel) [37]. The head group has number zero. Solid line: simulated result. Squares: experimental values. O.P. = order parameter,  $N$  = carbon atom number.

### 2.3.4. Atomic Properties: Diffusion in Solutions and in Membranes

The dynamic properties of simulations of molecular systems are hard to test by comparison with experiment. An exception is the diffusion constant. Its calculation and comparison with experiment is a standard test for models of liquid water. The diffusion constant of the simple three-point charge (SPC) water model is  $4.3 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$  compared with an experimental value of  $2.7 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$  at 305 K.<sup>[38]</sup> Inclusion of a correction term for self-polarization in the SPC model led to a reparametrization, to the SPC/E (extended simple point charge) model, which yields a considerably smaller diffusion constant of  $2.5 \times 10^{-5} \text{ cm}^2 \text{ s}^{-1}$ .<sup>[39]</sup>

In the bilayered sodium ion/water/decanoate/decanol system mentioned in the previous section the simulated diffusion constants are:  $2.7 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$  (decanoate) and  $5.2 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$  (decanol), which values are to be compared with the value of  $1.5 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$  measured with nitroxide spin labels.

### 2.3.5. System Properties: Thermodynamic Quantities

Instead of atomic properties, thermodynamic system properties like the density or free energy of solvation can be calculated from a simulation and compared to experimental values.

For a crystal of a cytidine derivative X-ray diffraction data and MD simulation data were compared at two different temperatures.<sup>[40]</sup> Upon raising the temperature from 113 K to 289 K the volume of the unit cell increased by 2.5% in the simulation at constant pressure, which value is to be compared with an experimentally determined increase of 3.7%. At 289 K the simulated density was only 1.3% too large. A slightly larger discrepancy was found in a MD simulation of crystalline cyclosporin A: the simulated density of  $1.080 \text{ g cm}^{-3}$  was 3.6% larger than the experimental value of  $1.042 \text{ g cm}^{-3}$ .<sup>[30]</sup> Comparable deviations have been found for other systems: the SPC/E model yields  $0.998 \text{ g cm}^{-3}$  at 306 K, compared with an experimental value of  $0.995 \text{ g cm}^{-3}$  at 305 K.<sup>[39]</sup>

Finally, we quote an example of a free energy of hydration. For the solvation of methanol in water the standard GROMOS parameters yield a free enthalpy of  $7 \text{ kJ mol}^{-1}$  compared with an experimental value of  $5 \text{ kJ mol}^{-1}$ .<sup>[28]</sup> However, when charged moieties are involved the accuracy by which free energy of solvation can be calculated is no better than about  $10\text{--}20 \text{ kJ mol}^{-1}$ .<sup>[28, 41]</sup>

## 2.4. Is Computer Simulation Useful?

The prediction of properties by computer simulation of complex molecular systems is certainly not accurate enough to justify abandoning the measurement of properties. If a measurement is not too difficult it is always to be preferred over a prediction by simulation. The utility of computer simulation studies does not lie in the (still remote) possibility of replacing experimental measurement, but rather in its ability to complement experiments. Quantities that are inaccessible to experiment can be monitored in computer simulations.

We consider the computer simulation of complex molecular systems to be useful for the following reasons.

1. It provides an *understanding of the relation* between *microscopic* properties and *macroscopic* behavior. In a computer a microscopic molecular model and force field can be changed at will and the consequences for the macroscopic behavior of the molecular system can be evaluated.
2. During the last few years computer simulation has become a standard tool in the *determination of spatial molecular structure* on the basis of X-ray diffraction or 2D-NMR data.
3. Under favorable conditions computer simulation can be used to obtain *quantitative estimates* of quantities like binding constants of ligands to receptors. This is especially useful where the creation of the ligand or the measurement of its binding constant is costly or time-consuming.
4. Finally we mention the possibility of carrying out simulation under extreme (unobservable) conditions of temperature and pressure.

### 3. Methodology

Here we briefly describe the methodology of classical computer simulation as it is applied to complex molecular systems.

#### 3.1. Choice of Molecular Model and Force Field

##### 3.1.1. Current Force Fields for Molecular Systems

A typical molecular force field or effective potential for a system of  $N$  atoms with masses  $m_i$  ( $i = 1, 2, \dots, N$ ) and cartesian position vectors  $\mathbf{r}_i$  has the form (1).

$$\begin{aligned}
 V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = & \sum_{\text{bonds}} \frac{1}{2} K_b [b - b_0]^2 + \sum_{\text{angles}} \frac{1}{2} K_\theta [\theta - \theta_0]^2 \\
 & + \sum_{\text{improper}} \frac{1}{2} K_\xi [\xi - \xi_0]^2 \\
 & + \sum_{\text{dihedrals}} K_\phi [1 + \cos(n\phi - \delta)] \\
 & + \sum_{\text{pairs}(i,j)} [C_{12}(i,j)/r_{ij}^{12} - C_6(i,j)/r_{ij}^6 \\
 & + q_i q_j / (4\pi \epsilon_0 \epsilon_r r_{ij})]
 \end{aligned} \quad (1)$$

The first term represents the covalent bond stretching interaction along bond  $b$ . It is a harmonic potential in which the minimum energy bond length  $b_0$  and the force constant  $K_b$  vary with the particular type of bond. The second term describes the bond angle bending (three-body) interaction in similar form. Two forms are used for the (four-body) dihedral angle interactions: a harmonic term for dihedral angles  $\xi$  that are not allowed to make transitions, e.g. dihedral angles within aromatic rings, and a sinusoidal term for the other dihedral angles  $\phi$ , which may make 360 degree turns. The last term is a sum over all pairs of atoms and represents the effective nonbonded interaction, composed of the van der Waals and the Coulomb interaction between atoms  $i$  and  $j$  with charges  $q_i$  and  $q_j$  at a distance  $r_{ij}$ .

There exists a large number of variants of expression (1).<sup>[11-13, 42-53]</sup> Some force fields contain mixed terms like

$K_{\theta b}[b - b_0][\theta - \theta_0]$ , which directly couple bond-length and bond-angle vibrations.<sup>[42]</sup> Others use more complex dihedral interactions terms.<sup>[43, 44]</sup> The choice of the relative dielectric constant  $\epsilon_r$  is also a matter of dispute. Values ranging from  $\epsilon_r = 1$ <sup>[11, 12]</sup> to  $\epsilon_r = 8$ <sup>[45]</sup> have been used, while others take  $\epsilon_r$  proportional to the distance  $r_{ij}$ .<sup>[22, 46, 47]</sup> Sometimes the Coulomb term is completely ignored.<sup>[44]</sup> Although hydrogen bonding can be appropriately modeled using expression (1),<sup>[11, 12, 48, 49]</sup> in some force fields special hydrogen bonding potential terms are used to ensure proper hydrogen bonding.<sup>[43-47]</sup> Another way to refine expression (1) is to allow for non-atomic interaction centers or virtual sites, that is, interactions between points (e.g. lone pairs) not located on atoms.<sup>[50]</sup> For solvents, especially water, a variety of molecular models is available,<sup>[19, 38, 51]</sup> of which a few have been developed explicitly for use in mixed solute-water systems.<sup>[11, 49]</sup> Models for nonpolar solvents like carbon tetrachloride are also available.<sup>[52]</sup>

When determining the parameters of the interaction function (1) there are essentially two routes to take. The most elegant procedure is to fit them to results (potential or field) of ab-initio quantum calculations on small molecular clusters. However, due to various serious approximations that have to be made in this type of procedure, the resulting force fields are in general not very satisfactory. The alternative is to fit the force field parameters to experimental data (crystal structure, energy and lattice dynamics, infrared, X-ray data on small molecules, liquid properties like density and enthalpy of vaporization, free energies of solvation, nuclear magnetic resonance data, etc.). In our opinion the best results have been obtained by this semi-empirical approach.<sup>[11, 53]</sup>

We wish to stress that one should fit force field parameters to properties of small molecules, which may be considered as building blocks of larger molecules such as proteins and DNA, and subsequently apply them to these larger molecules as a test without any further adaptations to improve the test results.

The choice of a particular force field should depend on the type of system for which it has been designed. The MM2 force field<sup>[13]</sup> is based on gas phase structures of small organic compounds. The AMBER<sup>[47]</sup> and CHARMM<sup>[46]</sup> force fields are aimed at a description of isolated polypeptides and polynucleotides, in which the absence of a solvent (aqueous) environment is compensated by the use of a distance-dependent dielectric constant  $\epsilon_r$ . The ECEPP<sup>[43, 54]</sup> and UNICEPP<sup>[55]</sup> force fields use  $\epsilon_r = 4$ , whereas the GRO-MOS<sup>[12, 49]</sup> force field uses  $\epsilon_r = 1$ , since it has been set up for simulation of biomolecules in aqueous environment. This also holds for the OPLS<sup>[11]</sup> force field which is aimed at a proper description of solvation properties. Force fields that are applicable to a more restricted range of compounds like ions,<sup>[56]</sup> liquid metals,<sup>[57]</sup> salts,<sup>[6]</sup> or carbohydrates,<sup>[58]</sup> are manifold. The quality of the various force fields should be judged from the literature concerning their application to molecular systems.

##### 3.1.2. Inclusion of Polarizability

The term in the interaction function (1) representing the nonbonded interactions consists only of a summation over

all pair interactions in the system. Nonbonded many-body interactions are neglected. Yet, inclusion of polarizability of atoms or bonds will be inevitable if one would like to simulate, e.g., the binding of a charged ligand, which will polarize the part of the receptor to which it is binding (Table 1). About 10% (4 kJ mol<sup>-1</sup>) of the energy of liquid water is polarization energy. The infinite frequency dielectric constant of water is  $\epsilon_\infty = 5.3$ , its effective dipole moment changes from 1.85 D in the gas phase to 2.4 D in the liquid. An analysis of the contribution of polarizability to local fields in proteins is given in Ref. [59].

Inclusion of polarizability in the molecular model is not too difficult, as can be observed from the following simple outline.<sup>[60]</sup> We consider a system of  $N$  point dipoles with cartesian position vectors  $\mathbf{r}_i$ , dipole moments  $\mathbf{p}_i$  and polarizabilities  $\alpha_i$  (constant, isotropic). The induced dipoles  $\Delta\mathbf{p}_i$  obey a field equation (2), where  $\mathbf{E}_i$  denotes the electric field

$$\Delta\mathbf{p}_i = \alpha_i \mathbf{E}_i = \alpha_i \sum_{\substack{j=1 \\ \neq i}}^N \mathbf{T}_{ij} [\mathbf{p}_j + \Delta\mathbf{p}_j] \quad (2)$$

at position  $\mathbf{r}_i$  and the field tensor  $\mathbf{T}_{ij}$  is given by Equation (3). The field equation can be solved for  $\Delta\mathbf{p}_i$  either by the inver-

$$\mathbf{T}_{ij} = (4\pi\epsilon_0)^{-1} [3 \mathbf{r}_i \mathbf{r}_j / r_{ij}^2 - 1] r_{ij}^{-3} \quad (3)$$

sion of a matrix of size  $3N \times 3N$ , or by iteration. The former method is impractical for a molecular system containing thousands of atoms. The latter method is well suited for use in MD simulations, since the induced dipoles at the previous MD integration time step,  $\Delta\mathbf{p}_i(t - \Delta t)$ , will be an excellent starting point for the iterative solution of the field equation (2) at time  $t$ , which yields  $\Delta\mathbf{p}_i(t)$ . In this way inclusion of polarizability may increase simulation times by only 20–100%.

Computer simulations that include polarizability have only been performed for a limited number of molecular systems, such as ionic liquids,<sup>[10, 61]</sup> water,<sup>[62]</sup> and a single protein.<sup>[63, 64]</sup> Many aspects of the treatment of polarizability in MD simulations are still under investigation. What is the most practical way to model polarizability?

- (1) By inducing point dipoles, as sketched above,
- (2) by changing the magnitudes of (atomic) charges,
- (3) by changing the positions of (atomic) charges.

How should the size of the atomic polarizabilities  $\alpha_i$  be chosen, etc?

### 3.1.3. Treatment of Long Range Coulomb Forces

The summation of the last term in the interaction function (1) covering the nonbonded interaction runs over all atom pairs in the molecular system. It is proportional to  $N^2$ , the square of the number of atoms in the system. Since the other parts of the calculation are proportional to  $N$ , computational efficiency can be much improved by a reduction of this summation. The simplest procedure is to apply a cut-off criterion for the nonbonded interaction and to use a list of neighbor atoms lying within the cut-off, which is only updated every so many simulation steps. The cut-off radius  $R_C$  usually has a value between 6 Å and 9 Å and the neighbor list is updated about every 10 or 20 MD time steps. This proce-

dures does not introduce any errors as long as the range of the nonbonded interaction is smaller than  $R_C$ . However, the Coulomb term in (1) is proportional to  $r^{-1}$ , which makes it long-ranged. If the molecular model does not involve bare (partial) charges on atoms, but only dipoles or higher multipoles, the electrostatic interaction term becomes proportional to  $r^{-3}$ , which makes it of much shorter range. However, when dipoles are correlated over larger distances, as is the case for secondary structure elements like  $\alpha$ -helices in proteins, their interactions again become long-ranged.<sup>[65]</sup>

In the following subsections we briefly discuss a variety of methods for the treatment of long-range electrostatic interactions in molecular systems.<sup>[66, 67]</sup>

#### 3.1.3.1. Distance Dependent Dielectric Constant

A simple way to reduce the range of the Coulomb interaction is to introduce a relative dielectric constant proportional to  $r$ , viz.  $\epsilon_r = r$  (in Å). The interaction becomes proportional to  $r^{-2}$ . It is difficult to find a physical argument in favor of this approximation. Its overall effect is to effectively reduce all types of long-ranged interactions. Due to its simplicity it has been incorporated into a number of current force fields.<sup>[46, 47]</sup> Nevertheless, we think this approximation is too crude for practical applications.

#### 3.1.3.2. Cut-off Radius and Neutral Groups of Atoms

When applying a cut-off radius  $R_C$ , the discontinuity of the interaction at a distance  $r = R_C$  will act as a noise source in a MD simulation, and this will artificially increase the kinetic energy of the atoms and thus the temperature of the system. A possible way to reduce the noise is to multiply the nonbonded interaction term in (1) with a so-called switching function (4), which satisfies the conditions  $S(R_S) = 1$ ,

$$S(r) = \begin{cases} 1 & r < R_S \\ (R_C - r)^2 (R_C + 2r - 3R_S) / (R_C - R_S)^3 & R_S < r < R_C \\ 0 & r > R_C \end{cases} \quad (4)$$

$dS/dr(R_S) = 0$ ,  $S(R_C) = 0$ ,  $dS/dr(R_C) = 0$ . Its effect is to smoothen the interaction on the interval  $(R_S, R_C)$ , but there is no physical argument for its use. An empirical evaluation of the use of switching functions can be found in Ref. [68].

When the (partial) atomic charges of a group of atoms add up to exactly zero, the leading term of the electric interaction between two such groups of atoms is of dipolar character, that is, proportional to  $r^{-3}$ . For larger  $r$  the sum of the  $r^{-1}$  monopole contributions of the various atom pairs to the group–group interaction will become zero. Therefore, the range of the electric interaction can be considerably reduced when atoms are assembled in neutral groups, so-called *charge groups*, which have a zero net charge, and for which the electric interaction with other (groups of) atoms is either calculated for all atoms of the charge group or for none.<sup>[12, 69]</sup> When using the charge group concept the cut-off criterion should be applied to the distance between groups, and a switching function like (4) must *not* be used, since it distorts in the interval  $(R_S, R_C)$  the proper  $r^{-1}$  weighting of atom–atom monopole interactions.

### 3.1.3.3. Cut-off Radius plus Multipole Expansion

The technique of using neutral groups of atoms is based on the more general fact that a charge distribution of a finite group of atoms can be approximated by a multipole expansion (monopole, dipole, quadrupole, etc.). The electric interaction between two groups of atoms can be formulated as the product of the two multipole expansions. The terms in the resulting expression can be grouped according to their distance dependence: monopole-monopole ( $r^{-1}$ ), monopole-dipole ( $r^{-2}$ ), monopole-quadrupole and dipole-dipole ( $r^{-3}$ ), etc. At long distance only the leading terms in the series need to be taken into account. Application of this multipole expansion approximation in MD simulations was suggested by Ladd.<sup>[70]</sup> It is also used in combination with the "cut-off plus atom pair list" technique in Ref. [46].

### 3.1.3.4. Twin Range Method

The twin range method<sup>[66]</sup> is illustrated in Figure 7. Two cut-off radii are used. The atoms  $j$  lying within a distance  $R_C^1$  from atom  $i$  are stored in a neighbor list of atom  $i$ . The interaction of the atoms  $j$  for which  $R_C^1 < r_{ij} < R_C^2$  with atom  $i$ , is stored in the form of a so-called long-range force  $F_i^{lr}$  on atom  $i$ . At each MD time step the nonbonded interaction consists of two contributions: (1) the short-range part which is calculated from the neighbor list using the actual atom positions, and (2) the long range part  $F^{lr}$  which is kept fixed during  $N_C^1$  time steps. Neighbor list and long range force  $F^{lr}$  are simultaneously updated every  $N_C^1$  (10–100) time steps.

This twin range method is based on the assumption that the high-frequency components of the long range force may be safely neglected. For example, the mean and low frequency field of the correlated peptide dipoles of the long  $\alpha$ -helix are accurately accounted for, only the fast ( $\leq 0.2$  ps) vibrations are neglected (see Fig. 7). The twin range method can also be applied using charged groups instead of atoms.<sup>[12]</sup>

### 3.1.3.5. Continuum Approximations to the Reaction Field

In the previous subsections approximations to the long-range interaction between atoms were discussed that are ex-

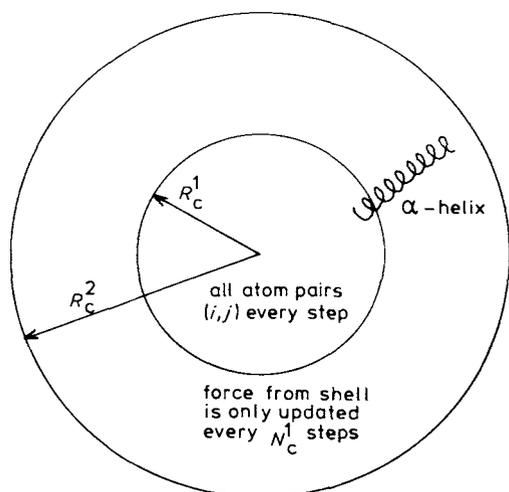


Fig. 7. Twin range method. High-frequency components of the force on the central atom exerted by atoms between  $R_C^1$  and  $R_C^2$  are neglected.

PLICITLY taken into account as degrees of freedom in the system that is simulated. If parts of the system are homogeneous, like the bulk solvent surrounding a solute, the number of atoms or degrees of freedom can be reduced considerably by modeling of the homogeneous part as a continuous medium, e.g. a continuous dielectric. In this type of approximation the system is divided into two parts: (1) an inner region where the atomic charges  $q_i$  are explicitly treated (dielectric

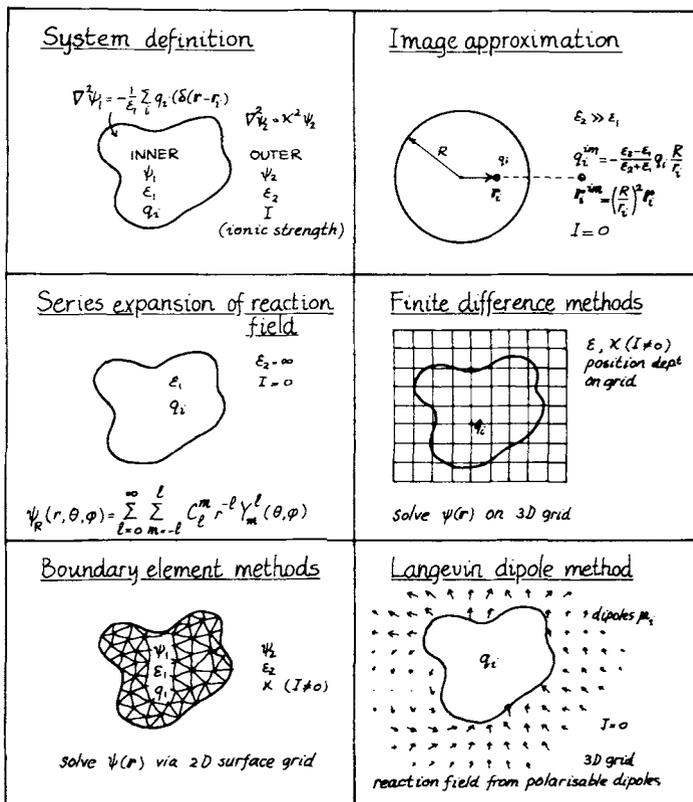


Fig. 8. Non-periodic methods for computing long-range Coulomb forces.

constant  $\epsilon_1$ ), and (2) an outer region which is treated as a continuous medium with dielectric constant  $\epsilon_2$  and ionic strength  $I$  (see Fig. 8). The potential in the inner region

$$\psi_1(r) = \psi_C(r) + \psi_R(r) \quad (5)$$

[Eq. (5)] consists of two terms. The first one [Eq. (6)] is a direct Coulomb term due to the charges in the inner region,

$$\psi_C(r) = (4\pi\epsilon_0\epsilon_1)^{-1} \sum_{i=1}^N q_i/|r-r_i| \quad (6)$$

which is a solution to the Poisson equation (7) for the inner

$$\nabla^2 \psi(r) = -\epsilon_1^{-1} \sum_{i=1}^N q_i \delta(r-r_i) \quad (7)$$

region. The second term is the reaction field potential  $\psi_R(r)$ , which satisfies the Poisson equation with zero source terms (no charges), that is, the Laplace equation (8). This means

$$\nabla^2 \psi(r) = 0. \quad (8)$$

that  $\psi_1(\mathbf{r})$  is also a solution of (7). The potential in the outer region is denoted by  $\psi_2(\mathbf{r})$ . If the outer region continuum has a finite ionic strength  $I$ , the potential  $\psi_2(\mathbf{r})$  must be a solution of the Poisson-Boltzmann equation (9) with inverse Debye

$$\nabla^2 \psi(\mathbf{r}) = \kappa^2 \psi(\mathbf{r}) \quad (9)$$

length  $\kappa = 2IF^2/(\varepsilon_2 RT)$  in which  $F$  is Faraday's constant,  $R$  is the gas constant, and  $T$  the temperature. For zero ionic strength, (9) reduces to (8). The boundary condition at infinity is given by (10), and at the boundary between the inner

$$\lim_{r \rightarrow \infty} \psi_2(\mathbf{r}) = 0 \quad (10)$$

and outer regions the potential  $\psi(\mathbf{r})$  [Eq. (11)] and the dielectric displacement  $\varepsilon \nabla \psi(\mathbf{r})$  [Eq. (12)] must be continuous:

$$\psi_1(\mathbf{r}) = \psi_2(\mathbf{r}) \quad \mathbf{r} = \text{boundary} \quad (11)$$

$$\varepsilon_1 \nabla_n \psi_1(\mathbf{r}) = \varepsilon_2 \nabla_n \psi_2(\mathbf{r}) \quad \mathbf{r} = \text{boundary} \quad (12)$$

where the component of the gradient normal to the boundary is denoted by  $\nabla_n$ .

The exact form of  $\psi(\mathbf{r})$  will depend on the shape of the boundary, the values of  $\varepsilon_2$  and  $\kappa$ , and the computational method that is used to solve the field equations (7)–(12).

### 3.1.3.6. Continuum Methods:

#### Image Charge Approximation to the Reaction Field

When the boundary is a sphere of radius  $R$  the reaction field due to a charge  $q_i$  located at position  $\mathbf{r}_i$  with respect to the origin of the sphere can be approximated (dipolar approximation,  $\varepsilon_2 \gg \varepsilon_1$ ) by the field that is generated by a so-called image charge  $q_i^{\text{im}} = -(\varepsilon_2 - \varepsilon_1)(\varepsilon_2 + \varepsilon_1)q_i R/r_i$  that is located at position  $\mathbf{r}_i^{\text{im}} = (R/r_i)^2 \mathbf{r}_i$  [71] (see Fig. 8). The name of this approximation is due to the mirror type of behavior of the spherical boundary: if  $q_i$  approaches the boundary,  $q_i^{\text{im}}$  will do likewise, causing a pole in the potential at the boundary. This means that the image approximation breaks down for charges close to the boundary. The other limitations of the method are the requirement of a spherical boundary, and the conditions that  $I = 0$  and  $\varepsilon_2 \gg \varepsilon_1$ .

### 3.1.3.7. Continuum Methods:

#### Series Expansion of the Reaction Field

When the boundary is of irregular shape the reaction field potential cannot be written in a closed form, but must be approximated numerically. If we assume that  $I = 0$  and  $\varepsilon_2 = \infty$ , the following approach is possible. Since the reaction field potential  $\psi_R(\mathbf{r})$  must satisfy the Laplace equation (8) it may be expanded in Legendre polynomials, which also satisfy Equation (8) [Eq. (13)]. If  $\varepsilon_2 = \infty$ , we have  $\psi_2(\mathbf{r}) = 0$

$$\psi_R(\mathbf{r}, \theta, \varphi) = \sum_{l=0}^{\infty} \sum_{m=-l}^{+l} C_l^m r^{-l} Y_l^m(\theta, \varphi). \quad (13)$$

at the boundary, or using (11) and (5) Equation (14), at the boundary, with  $\psi_c(\mathbf{r})$  fixed by (6). The expansion coefficients

$$\psi_R(\mathbf{r}) = -\psi_c(\mathbf{r}) \quad (14)$$

$C_l^m$  can be determined numerically by performing a least squares fit of  $\psi_R(\mathbf{r})$  to the given  $-\psi_c(\mathbf{r})$  at a chosen number of points on the boundary.

The method works (Berendsen and Zwinderman, private communication), but is rather expensive due to the matrix inversion inherent in the least squares fit procedure. Close to the boundary the approximation is poor. The other limitations are that  $I = 0$  and  $\varepsilon_2 = \infty$ .

### 3.1.3.8. Continuum Methods:

#### Three-dimensional Finite Difference Techniques

Another numerical approach is to compute the electric potential on a three-dimensional (3D) grid using finite difference methods (Fig. 8). The atomic charges  $q_i$  are distributed over grid points. To each grid point a position-dependent dielectric constant  $\varepsilon_1$  is assigned. Then the field equations (7)–(12) can be numerically solved.<sup>[72]</sup> This involves the solution of a set of linear equations for the potential at grid points. The method is too time-consuming to be used in MD simulations, since the density of grid points must be sufficient to adequately represent the atomic charge distribution. The use of a position-dependent dielectric constant introduces an arbitrary element into the method.

### 3.1.3.9. Continuum Methods:

#### Two-dimensional Boundary Element Techniques

A recent development is the application of Green's function techniques to the electric field problem.<sup>[73]</sup> The Green's

$$G_1(\mathbf{r}|\mathbf{r}_0) = \frac{1}{4\pi|\mathbf{r} - \mathbf{r}_0|} \quad (15)$$

function (15) for the inner region satisfies Equation (16) (a Poisson equation with charge  $\varepsilon_0$  at  $\mathbf{r}_0$ ). The Green's func-

$$\nabla^2 G_1(\mathbf{r}|\mathbf{r}_0) = -\delta(\mathbf{r} - \mathbf{r}_0) \quad (16)$$

tion for the outer region [Eq. (17)] satisfies Equation (18) (a Poisson-Boltzmann equation with charge  $\varepsilon_0$  at  $\mathbf{r}_0$ ).

$$G_2(\mathbf{r}|\mathbf{r}_0) = \frac{e^{-\kappa|\mathbf{r} - \mathbf{r}_0|}}{4\pi|\mathbf{r} - \mathbf{r}_0|} \quad (17)$$

$$\nabla^2 G(\mathbf{r}|\mathbf{r}_0) = \kappa^2 G(\mathbf{r}|\mathbf{r}_0) - \delta(\mathbf{r} - \mathbf{r}_0) \quad (18)$$

The next step is to integrate the following integrand over the

$$[G_1 \times \text{expression (7) for } \psi_1 - \psi_1 \times \text{expression (16) for } G_1] \quad (19)$$

inner region and to convert it by Green's theorem into an integral over the surface of the boundary region. This yields Equation (20), consisting of a surface term and a source term.

$$\psi_1(\mathbf{r}) = \int_{\text{surface}} [G_1(\mathbf{r}'|\mathbf{r}) \nabla_n \psi_1(\mathbf{r}') - \psi_1(\mathbf{r}') \nabla_n G_1(\mathbf{r}'|\mathbf{r})] d\sigma + \sum_{i=1}^N \frac{q_i}{4\pi\varepsilon_0|\mathbf{r} - \mathbf{r}_i|} \quad (20)$$

By integrating (21) and conversion we obtain equation (22).

$[G_2 \times \text{expression (9) for } \psi_2 - \psi_2 \times \text{expression (18) for } G_2]$  (21)

$$\psi_2(\mathbf{r}) = \int_{\text{surface}} [G_2(\mathbf{r}'|\mathbf{r}) \nabla_n \psi_2(\mathbf{r}') - \psi_2(\mathbf{r}') \nabla_n G_2(\mathbf{r}'|\mathbf{r})] d\sigma \quad (22)$$

Application of the boundary conditions (11) and (12) yields two linear integral equations in  $\psi_1$  and  $\nabla_n \psi_1$  on the boundary surface. These equations can be solved numerically by discretization on a surface net, and subsequently solved by matrix inversion. The (large) matrix that is to be inverted only depends on the shape and position of the boundary surface, *not* on the positions of the charges in the inner region. This means that in a MD simulation the time-consuming matrix inversion can be avoided as long as the boundary remains fixed.

The promise of this method lies in the reduction of a 3D problem to a 2D problem. On the boundary surface a charge density is generated which produces the proper reaction field. Also the methods of *Shaw*,<sup>[74]</sup> and of *Zauhar and Morgan*<sup>[75]</sup> transform the problem from a three-dimensional one into a two-dimensional one.

### 3.1.3.10. Langevin Dipole Model

The Langevin dipole model of *Warshe*<sup>[63, 67]</sup> uses a 3D-grid, but it is not a continuum method. The medium in the outer region is mimicked by polarizable point dipoles  $\mu_i$  on the grid points  $i$  of a 3D grid. The dipoles may be thought of as representing the molecular dipoles of the solvent molecules forming the outer region, so the spacing of the grid points corresponds to the size of the solvent molecules. The size and direction of a dipole  $\mu_i$  is determined by the electric field  $E_i$  at the grid point according to the Langevin formula [Eq. (23)]. Here,  $\mu_0$  is the magnitude of the dipole moment

$$\mu_i = \mu_0 \left[ \frac{e^{+C\mu_0 E_i/kT} + e^{-C\mu_0 E_i/kT}}{e^{+C\mu_0 E_i/kT} - e^{-C\mu_0 E_i/kT}} - \frac{1}{C\mu_0 E_i/kT} \right] \frac{E_i}{E_i} \quad (23)$$

of a solvent molecule, and  $C$  is a parameter representing the molecular resistance to reorientation. The model has been applied to protein simulations.<sup>[64]</sup> Questionable features of this model are the representation of the field in the outer region and the discreteness near the boundary.

### 3.1.3.11. Periodic Lattice Summation Methods

In computer simulations of liquids or solutions, periodic boundary conditions are often used to minimize the boundary effects. The computational box containing the molecular system is surrounded by an infinite number of copies of itself (see Fig. 9). In this way an infinite periodic system is simulated.

The electric interaction in a periodic system is obtained by a summation over all atom pairs in the central box (Fig. 9) and over all atom pairs of which one atom lies in the central box and the other is a periodic image [Eq. (24)].

$$E = (4\pi\epsilon_0)^{-1} \frac{1}{2} \sum_{n=0}^{\infty} \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}|} \quad (24)$$

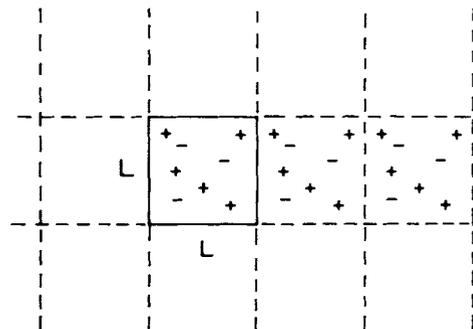


Fig. 9. Periodic cubic system. Central computational box (solid line) with two of the infinite number of periodic image boxes.

The sum over  $\mathbf{n}$  is a summation over all simple cubic lattice points  $\mathbf{n} = (n_x L, n_y L, n_z L)$  with  $n_x, n_y, n_z$  integers. The infinite sum of point charges must be converted into a form which converges faster than Equation (24).

### 3.1.3.12. Periodic Methods: the Ewald Sum

The Ewald sum is a technique for computing the interaction of a charge and all its periodic images.<sup>[6, 76-78]</sup> The charge distribution  $\rho(\mathbf{r})$  in the system is an infinite set of point charges, mathematically represented as delta functions [Eq. (25)].

$$\rho_i(\mathbf{r}) = q_i \delta(\mathbf{r} - \mathbf{r}_i) \quad (25)$$

Each point charge  $q_i$  is now surrounded by an isotropic Gaussian charge distribution of equal magnitude and opposite sign [Eq. (26)].

$$\rho_i^G(\mathbf{r}) = -q_i (\alpha/\sqrt{\pi})^3 e^{-\alpha^2 |\mathbf{r} - \mathbf{r}_i|^2} \quad (26)$$

This smeared charge screens the interaction between the point charges, so that the interaction calculated using the screened charge distribution (27) becomes short-ranged [Eq. (28)] due to the appearance of the complementary error function (29).

$$\rho_i^S(\mathbf{r}) \equiv \rho_i(\mathbf{r}) + \rho_i^G(\mathbf{r}) \quad (27)$$

$$E^S = (4\pi\epsilon_0)^{-1} \frac{1}{2} \sum_{n=0}^{\infty} \sum_{i=1}^N \sum_{j=1}^N \frac{q_i q_j}{|\mathbf{r}_{ij} + \mathbf{n}|} \operatorname{erfc}(\alpha |\mathbf{r}_{ij} + \mathbf{n}|) \quad (28)$$

$$\operatorname{erfc}(x) = 2\pi^{-1/2} \int_x^{\infty} e^{-y^2} dy \quad (29)$$

Thus,  $E^S$  can be well approximated using a finite summation in (28). Of course a purely Gaussian charge distribution  $-\rho_i^G(\mathbf{r})$  must be added to  $\rho_i^S(\mathbf{r})$  in order to recover the original charge distribution  $\rho_i(\mathbf{r})$ . The interaction of these Gaussian distributions is expressed as a lattice sum in reciprocal space minus a self term [Eq. (30)], where

$$E^G = (4\pi\epsilon_0)^{-1} 2\pi L^{-3} \sum_{k \neq 0}^{\infty} k^{-2} e^{-k^2/(4\alpha^2)} \left| \sum_{j=1}^N q_j e^{-ik \cdot \mathbf{r}_j} \right|^2 - (4\pi\epsilon_0)^{-1} \pi^{-1/2} \alpha \sum_{j=1}^N q_j^2 \quad (30)$$

$\mathbf{k} = 2\pi L^{-1}(l_x, l_y, l_z)$  and  $l_x, l_y, l_z$  are integers. Due to the presence of the exponential factor the infinite lattice sum can be well approximated by a finite one. The parameter  $\alpha$  should be chosen such as to optimize the convergence properties of both sums (28) and (30) which contribute to  $E = E^s + E^G$ .

The Ewald sum technique is routinely used in simulations of ionic systems. When applied to non-crystalline systems such as liquids or solutions it has the disadvantage of enhancing the artifact of the application of periodic boundary conditions.

### 3.1.3.13. Periodic Methods: Fourier Techniques

The Poisson equation (31) for the potential  $\psi(\mathbf{r})$  is a sec-

$$\nabla^2 \psi(\mathbf{r}) = -\varepsilon_0^{-1} \varrho(\mathbf{r}) \quad (31)$$

ond-order partial differential equation in  $\mathbf{r}$ -space. However, when transformed into reciprocal space ( $\mathbf{k}$ -space) it becomes a simple algebraic equation (32) with the solution (33).

$$-k^2 \hat{\psi}(\mathbf{k}) = -\varepsilon_0^{-1} \hat{\varrho}(\mathbf{k}) \quad (32)$$

$$\hat{\psi}(\mathbf{k}) = (\varepsilon_0 k^2)^{-1} \hat{\varrho}(\mathbf{k}) \quad (33)$$

Here, the 3D Fourier transformation is defined according to Equation (34), and likewise for the charge density  $\varrho(\mathbf{r})$ . The inverse transform of Equation (33) yields (35).

$$\hat{\psi}(\mathbf{k}) \equiv F\{\psi(\mathbf{r})\} \equiv (2\pi)^{-3/2} \int \psi(\mathbf{r}) e^{-i\mathbf{k} \cdot \mathbf{r}} d\mathbf{r} \quad (34)$$

$$\psi(\mathbf{r}) = F^{-1}\{(\varepsilon_0 k^2)^{-1} F\{\varrho(\mathbf{r})\}\} \quad (35)$$

Due to the availability of fast Fourier transform techniques Equation (35) is a fast method to solve the Poisson equation on a periodic three-dimensional grid. For a practical implementation of the method we refer to Ref. [1]. It has been employed in the simulation of salts.<sup>[11]</sup>

### 3.1.4. Summary

The choice of molecular model and force field is essential to a proper prediction of the properties of a system. Therefore, it is of great importance to be aware of the fundamental assumptions, simplifications and approximations that are implicit in the various types of models used in the literature. When treating molecular systems in which Coulomb forces play a role one should be aware of the way these long-ranged forces are handled. Since these forces play a dominant role in many molecular systems we have tried to give an overview of the different methods—and so the different approximations—that are in use.

From Table 1 and the discussions in the previous sections it should be clear that a “best” force field does *not* exist. What will be the best choice of model and force field will depend on the type of molecular system and the type of property one is interested in. This means that the molecular modeler must have a picture of the strengths and weaknesses of the variety of force fields that are available in order to make a proper choice.

## 3.2. Searching Configuration Space and Generating an Ensemble

Once the molecular model and force field  $V(\mathbf{r})$  have been chosen, a method to search configuration space for configurations with low energy  $V(\mathbf{r})$  has to be selected. Various methods are available, each with their particular strength and weakness, which depend on

- the form and type of the interaction energy function  $V(\mathbf{r})$ ,
- the number of degrees of freedom (size of the system),
- the type of degrees of freedom, viz. cartesian coordinates versus other coordinates (e.g. bond lengths, bond angles, torsional angles plus center of mass coordinates of a molecule).

### 3.2.1. Systematic Search Methods

If the molecular system contains only a small number of degrees of freedom (coordinates) and if  $V(\mathbf{r})$  does not have too many (relevant) minima upon variation of the degrees of freedom, it is possible to systematically scan the complete configuration space of the system. For example, one can describe conformations of  $n$ -alkanes in terms of C–C–C torsional angles, each of which has three conformational minima, *trans* ( $0^\circ$ ), *gauche*<sup>+</sup> ( $120^\circ$ ) and *gauche*<sup>−</sup> ( $-120^\circ$ ). To find the lowest energy conformation of  $n$ -decane (seven torsional angles) one would have to compute  $V(\mathbf{r})$  for  $3^7 = 2187$  combinations of torsional angles. The relative weight of the different conformations in the ensemble representing this molecule at temperature  $T$  is given by the Boltzmann factor (36).

$$e^{-V(\mathbf{r})/k_B T} \quad (36)$$

The computing effort required by a systematic search of the degrees of freedom of a system grows exponentially with their number. Only very small molecular systems can be treated by systematic search methods.<sup>[79, 80]</sup> The number of degrees of freedom that still can be handled within a reasonable computing time strongly depends on the complexity of the function  $V(\mathbf{r})$ , that is, the time required to compute  $V(\mathbf{r})$  for each configuration. A possibility for speeding up the calculation is to split the selection of low  $V(\mathbf{r})$  configurations into different stages. In a first stage a simplified low computing cost energy function  $V_{\text{simple}}(\mathbf{r})$  is used to quickly discard high  $V_{\text{simple}}(\mathbf{r})$  configurations. In a second stage the complete  $V(\mathbf{r})$  is only to be evaluated for the remaining configurations. This type of filtering procedure has been used to predict the loop structure in proteins<sup>[81]</sup> and to predict the stable conformation of small peptides (five amino acid residues),<sup>[82]</sup> for example. The basic problem of filtering using simplified forms of  $V(\mathbf{r})$  is to ensure that the simplified  $V_{\text{simple}}(\mathbf{r})$  is a correct projection of the complete function  $V(\mathbf{r})$ : when  $V_{\text{simple}}(\mathbf{r})$  is large,  $V(\mathbf{r})$  should also be large, otherwise a configuration with low energy  $V(\mathbf{r})$  might be discarded in the first stage due to a high energy  $V_{\text{simple}}(\mathbf{r})$ .

### 3.2.2. Random Search Methods

If a system contains too many degrees of freedom, straightforward scanning of the complete configuration space is impossible. In that case a collection of configura-

tions can be generated by random sampling. Such a collection becomes an ensemble of configurations when each configuration is given its Boltzmann factor [cf. Eq. (36)] as weight factor in the collection. Two types of random search methods can be distinguished:

- Monte Carlo methods, in which a sequence of configurations is generated by an algorithm which ensures that the occurrence of configurations is proportional to their Boltzmann factors [Eq. (36)].
- Other methods which produce an arbitrary (non-Boltzmann) collection of configurations, which can be transformed into a Boltzmann ensemble by applying the weight factors from Equation (36).

### 3.2.2.1. Monte Carlo Simulation

The Monte Carlo (MC) simulation procedure by which a (canonical) ensemble is produced consists of the following steps.

1. Given a starting configuration  $\mathbf{r}_s$  a new configuration  $\mathbf{r}_{s+1} = \mathbf{r}_s + \Delta\mathbf{r}$  is generated by random displacement of one (or more) atoms. The random displacements  $\Delta\mathbf{r}$  should be such that in the limit of a large number of successive displacements the available cartesian space of all atoms is uniformly sampled. This does not mean that the actual sampling must be carried out in cartesian space. It can be done, e.g., in internal coordinate space ( $r, \theta, \varphi$ ), but since the equivalent volume element is  $r^2 \sin \theta \, dr d\theta d\varphi$ , the sampling in internal coordinate space must be non-uniform in order to produce a uniform sampling in terms of cartesian coordinates.
2. The newly generated configuration  $\mathbf{r}_{s+1}$  is either accepted or rejected on the basis of an energy criterion involving the change  $\Delta E = V(\mathbf{r}_{s+1}) - V(\mathbf{r}_s)$  of the potential energy with respect to the previous configuration. The new configuration is accepted when  $\Delta E \leq 0$ , or if  $\Delta E > 0$  when  $e^{-\Delta E/k_B T} > R$ , where  $R$  is a random number taken from a uniform distribution over the interval (0,1).

Upon acceptance, the new configuration is counted and used as a starting point for the next random displacement. If the criteria are not met, the new configuration  $\mathbf{r}_{s+1}$  is rejected. This implies that the previous configuration  $\mathbf{r}_s$  is counted again and used as a starting point for another random displacement.

It is relatively easy to understand that this procedure will generate a Boltzmann ensemble. We consider two configurations  $\mathbf{r}_1$  and  $\mathbf{r}_2$  with energies  $E_1 = V(\mathbf{r}_1) < V(\mathbf{r}_2) = E_2$ . The probability of stepping from configuration  $\mathbf{r}_2$  to  $\mathbf{r}_1$  equals 1, the reverse step has a probability  $\exp(-(E_2 - E_1)/k_B T)$ . When the populations  $p_1$  and  $p_2$  of the two configurations are in equilibrium, one has detailed balance conditions (37) or (38).

$$p_2 \cdot 1 = p_1 \cdot e^{-(E_2 - E_1)/k_B T} \quad (37)$$

$$p_1/p_2 = e^{-E_1/k_B T}/e^{-E_2/k_B T} \quad (38)$$

Each configuration occurs with a probability proportional to its Boltzmann factor [Eq. (36)].

The advantage of this (Metropolis) Monte Carlo or Boltzmann sampling over random sampling is that most sampled

configurations are relevant (low energy), while with random sampling much computational effort is likely to be spent on irrelevant (high energy) configurations. In order to obtain high computational efficiency, one would like to combine a large (random) step size with a high acceptance ratio. This is possible when applying MC techniques to simulate simple atomic or molecular liquids.<sup>[6]</sup> However, for complex systems involving many covalently bound atoms, a reasonable acceptance ratio can only be obtained for very small step size.<sup>[83]</sup> This is due to the fact that a random displacement will inevitably generate a very high bond energy of the bonds of the displaced atom. This makes MC methods rather inefficient for (macro)molecular systems.

### 3.2.2.2. Distance Geometry Methods

Distance geometry (DG) is a general method for converting a set of bounds on distances between atoms into a configuration of these atoms that is consistent with these bounds. The emergence of 2D-NMR techniques has spurred a renewed interest in techniques to obtain three-dimensional molecular structures from atom-atom distance information.<sup>[84-86]</sup> The DG method is also applied in pharmacophore modeling<sup>[87]</sup> or enzyme substrate docking [*J. M. Blaney*, private communication] to generate a collection of ligand structures compatible with a set of atom-atom distance bounds.

In distance geometry a molecular structure is described in terms of the set of all pairwise interatomic distances, which can be written in the form of a so-called (symmetric) distance matrix. By entering maximum distances between atoms of pairs in the upper right-hand triangle, and minimum distances in the lower left-hand triangle of the distance matrix it becomes a distance bounds matrix. Such a matrix describes the complete configuration space accessible to the molecule within the specified bounds. In a distance geometry calculation a set of random configurations is generated by choosing atom-atom distances at random within the specified bounds, and subsequently converting the resulting distance matrix into a structure in three-dimensional cartesian space using a so-called embedding algorithm.<sup>[84-86]</sup>

The DG method is a powerful method for generating a set of configurations compatible with a set of atom-atom distances, but also has a number of limitations. It is not possible to apply an energy function  $V(\mathbf{r})$  like Equation (1) in a DG calculation. The energy function has to be converted into a function of atom-atom distances only, and subsequently it must be simplified to a set of bounds on these distances, by which procedure much of the information present in  $V(\mathbf{r})$  will be lost. As a consequence the DG method cannot properly handle solvent configurations, since a limited-distance description of a liquid lacks the ability to give proper statistical weight to the great variety of possible configurations. Another problem is the characterization of the distribution of generated three-dimensional configurations. Since the conversion from distance space into three-dimensional cartesian space is non-linear, a uniform sampling of distances between the lower and upper bounds will certainly not produce a set of configurations uniformly distributed in cartesian space.

### 3.2.2.3. Other Random Search Techniques

There exist infinitely many ways for generating a set of molecular configurations in a random manner. Whether the set of generated configurations may be considered to form an ensemble that can be used for statistical mechanical evaluation of quantities will depend on the sampling properties of the method that is used. It should sample the important (low energy) regions of configuration space and the configurations should occur according to their Boltzmann weight factors. Otherwise the set of configurations can only be viewed as such, not as an ensemble representative of the state of the system that is considered.

### 3.2.3. Dynamic Simulation Methods

Another way to generate an ensemble of configurations is to apply Nature's laws of motion for the atoms of a molecular system. This has the additional advantage that dynamical information about the system is obtained as well. The two major simulation techniques of this type are Molecular Dynamics (MD), in which Newton's equations of motion are integrated over time, and Stochastic Dynamics (SD), in which the Langevin equation of motion for Brownian motion is integrated over time.

#### 3.2.3.1. Molecular Dynamics Simulation

In the Molecular Dynamics (MD) method a trajectory (configurations as a function of time) of the molecular system is generated by simultaneous integration of Newton's equations of motion (39) and (40) for all the atoms in the

$$d^2\mathbf{r}_i(t)/dt^2 = m_i^{-1} \mathbf{F}_i \quad (39)$$

$$\mathbf{F}_i = -\partial V(\mathbf{r}_1, \dots, \mathbf{r}_N)/\partial \mathbf{r}_i \quad (40)$$

system. The force on atom  $i$  is denoted by  $\mathbf{F}_i$  and time is denoted by  $t$ . MD simulation requires calculation of the gradient of the potential energy  $V(\mathbf{r})$ , which therefore must be a differentiable function of the atomic coordinates  $\mathbf{r}_i$ . The integration of Equation (39) is performed in small time steps  $\Delta t$ , typically 1–10 fs for molecular systems. Static equilibrium quantities can be obtained by averaging over the trajectory, which must be of sufficient length to form a representative ensemble of the state of the system. In addition, dynamic information can be extracted. Another asset of MD simulation is that non-equilibrium properties can be efficiently studied by keeping the system in a steady non-equilibrium state, as is discussed in Section 5.4. Viewed as a technique to search configuration space, the power of MD lies in the fact that the kinetic energy present in the system allows it to surmount energy barriers that are of order of  $k_B T$  per degree of freedom. By raising the temperature  $T$  a larger part of conformation space can be searched, as has been shown by *DiNola et al.*,<sup>[88]</sup> who generated a series of different conformations of the hormone somatostatin by applying MD at  $T = 600$  K and at 1200 K. At the elevated temperature the total energy and potential energy are monitored for conspicuous fluctuations which may signal a possibly significant conformational change. When minima in the total energy occur, the system is cooled down and equilibrated at normal

temperature (300 K). In this way different conformations with comparable free energy were obtained. We note however that the search at elevated temperature favors selection of higher entropy conformations.

Searching conformation space by MD is expected to be efficient for molecules up to about 100 atoms. For larger molecules, which may and are likely to show a particular topological fold, MD methods will not be able to generate major topological rearrangements. Even when the barriers separating two topologically different low energy regions of conformation space are of the order of  $k_B T$ , the time needed for traversing them may be much too long to be covered in a MD simulation of 10–100 ps.

#### 3.2.3.2. Stochastic Dynamics Simulation

The method of stochastic dynamics (SD) is an extension of MD. A trajectory of the molecular system is generated by integration of the stochastic Langevin equation of motion (41).

$$d^2\mathbf{r}_i(t)/dt^2 = m_i^{-1} \mathbf{F}_i + m_i^{-1} \mathbf{R}_i - \gamma_i d\mathbf{r}_i(t)/dt \quad (41)$$

Two terms are added to Equation (39), a stochastic force  $\mathbf{R}_i$  and a frictional force proportional to a friction coefficient  $\gamma_i$ . The stochastic term introduces energy, the frictional term removes (kinetic) energy from the system, the condition for zero energy loss being that given in Equation (42), where

$$\langle \mathbf{R}_i^2 \rangle = 6 m_i \gamma_i k_B T_{ref} \quad (42)$$

$T_{ref}$  is the reference temperature of the system.

SD can be applied to establish a coupling of the individual atom motion to a heat bath,<sup>[89]</sup> or to mimic a solvent effect.<sup>[90, 91]</sup> In the latter case the stochastic term represents collisions of solute atoms with solvent molecules and the frictional term represents the drag exerted by the solvent on the solute atom motion. An introduction to SD simulation techniques is given in Refs. [90, 92].

### 3.2.4. Other Search Methods

There exists a variety of other methods for searching configuration space which are variants of the basic types discussed in the previous subsections.

The method of simulated annealing<sup>[93]</sup> is a MC or a MD simulation in which the temperature is gradually lowered to 0 K. In this way the system generally ends up in a lower energy state than when an ordinary gradient energy minimizer is used.

In a MD simulation the atomic velocities are large when the system explores the low energy regions of the potential energy function  $V(\mathbf{r})$ , and small when it crosses barriers at higher energy. For an efficient searching of configuration space one would like to invert this behavior: the atoms should move slowly in the valleys of  $V(\mathbf{r})$ . A search algorithm developed along this line seems to perform well for certain systems.<sup>[94]</sup> Within the framework of MD the ability of atoms to cross barriers can be greatly enhanced by keeping their kinetic energy nearly constant by a tight coupling to a heat bath.

Another possibility is the combination of different techniques; Monte Carlo simulation, gradient minimization, etc. in one computational scheme.<sup>[95, 96]</sup>

### 3.2.5. Summary

The various methods for searching configuration space can be classified as follows:

- A. Systematic search methods, which scan the complete configuration space of the molecular system.
- B. Methods which aim at generating a representative set of configurations. These may be divided into two types.
  1. Non-step methods, such as the DG method, which generate a (at least in principle) uncorrelated series of random configurations.
  2. Step methods, such as MC, MD and SD, which generate a new configuration from the previous one.

Most search techniques fall in this class. They can be distinguished by the way the step direction and the step size are chosen:

- according to the gradient  $-\nabla V(\mathbf{r})$ ,
- according to a memory of the path followed so far,  
or
- at random.

For example, in the MC method, the step direction is chosen at random, the actual step size is determined by the change in energy  $\Delta E$  (gradient) and a random element in case  $\Delta E > 0$ . In a MD simulation both step size and direction are determined by the force (gradient) and the velocity (memory). In general a search algorithm will make a step which is a linear combination of the gradient, previous steps (memory) and a random contribution. It will depend on the energy surface  $V(\mathbf{r})$  which is the optimal linear combination of these three ingredients.

Another classification of algorithms which generate a set of configurations is whether they produce an ensemble or not. Of the schemes discussed above only MC, MD and SD generate a Boltzmann weighted ensemble.

## 3.3. Boundary Conditions

When simulating a system of finite size, some thought must be given to the way the boundary of the system will be treated. The simplest choice is the vacuum boundary condition. When simulating a liquid, solution or solid rather than a molecule in the gas phase, it is common practice to minimize edge or wall effects by the application of periodic boundary conditions. If the irregularity of the system is incompatible with periodicity, edge or wall effects may be reduced by treating part of the system as an extended wall region in which the motion of the atoms is partially restricted.

### 3.3.1. Vacuum Boundary Condition

Simulation of a molecular system in vacuo, that is, without any wall or boundary, corresponds to the gas phase at zero pressure. When the vacuum boundary is used for a solid or a molecule in solution, properties of atoms near or at the surface of the system will be distorted.<sup>[97]</sup> The vacuum boundary condition may also distort the shape of a (non-

spherical) molecule, since it generally tends to minimize the surface area. Moreover, the shielding effect of a solvent with high dielectric permittivity, like water, on the electric interaction between charges or dipoles in a molecule is lacking in vacuo. We note that the water molecules do *not* need to be positioned between the charges or dipoles in order to produce a screening of the interaction between these. Therefore, simulation of a charged extended molecule like DNA in vacuo is a precarious undertaking. Solvation of DNA in a sphere with solvent molecules will shift the boundary effects from the DNA–vacuum interface to the water–vacuum interface and so improve the treatment of the DNA.<sup>[98]</sup> The best results in vacuo are obtained for relatively large globular molecular systems.

### 3.3.2. Periodic Boundary Conditions

The classical way to minimize edge effects in a finite system is to use periodic boundary conditions. The atoms of the system that is to be simulated are put into a cubic, or more generally into any periodically space-filling shaped box, which is treated as if it is surrounded by 26 ( $= 3^3 - 1^3$ ) identical translated (over distances  $\pm R_{\text{box}}$  in the  $x$ -,  $y$ -,  $z$ -directions) images of itself. The next layer of neighbor images of the central computational box contains  $5^3 - 3^3 - 1^3 = 98$  boxes, and so on. When an infinite lattice sum of the atomic interactions is to be performed (Sections 3.1.3.11–13), the interactions of an atom in the central computational box with *all* its periodic images are computed. In most cases this is not desirable. Then only interactions with nearest neighbors are taken into account. The black atom in the central computational box in Figure 10a will only interact with atoms or images of atoms that lie within the dashed line (nearest image, NI, or minimum image, MI, approximation). The anisotropy of the interaction due to the cubic shape of the nearest image box can be avoided by the application of a spherical cut-off (radius  $R_C$ ). The periodic boundary condition affects not only the computation of the forces, but also the positions of the atoms. It is common practice (though not necessary) to keep the atoms together, that is, in the central computational box: when an atom leaves the central box on one side, it enters it with identical velocity at the opposite side at the translated image position.

Application of periodic boundary conditions means that in fact a crystal is simulated. For a liquid or solution the periodicity is an artifact of the computation, so the effects should be minimized. An atom should not simultaneously interact with another atom and a periodic image of that atom. Consequently the length  $R_{\text{box}}$  of the edge of the periodic box should exceed twice the cut-off radius  $R_C$ . Possible distorting effects of the periodic boundary condition<sup>[99–102]</sup> may be traced by simulation of a system in differently shaped boxes (see below) of different size.

When simulating a spherical solute, use of a more spherically shaped computational box instead of a cubic or rectangular one may considerably reduce the number of solvent molecules that is needed to fill the remaining (after insertion of the solute) empty space in the box. A more spherically shaped space-filling periodic box is a truncated octahedron, (Fig. 10b).<sup>[103]</sup> It is obtained by cutting off the corners of a cube in such a way that the symmetry of the cube is main-

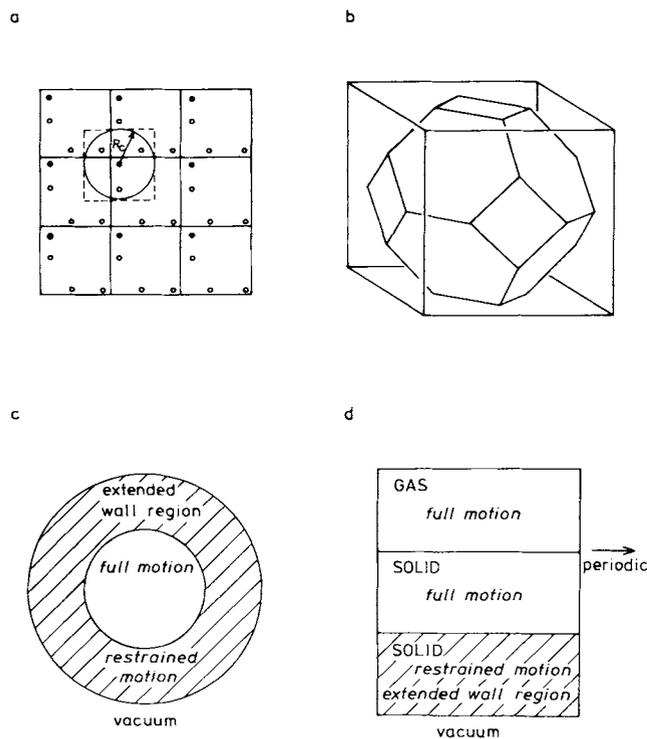


Fig. 10. Types of boundary conditions: a) cubic periodic, b) truncated octahedron periodic, c) spherical extended wall region, d) flat extended wall region for a gas-solid simulation.

tained, and such that the distance between opposite hexagonal planes equals  $\frac{1}{2}\sqrt{3}$  times the distance between opposite square planes. In this way half the volume of the original cube is cut away. The ratio of the inscribed-sphere volume to total volume of the truncated octahedron is 0.68, whereas it is only 0.52 for a cube. Use of a truncated octahedron instead of a rectangular periodic box may yield a sizeable reduction of the system to be simulated. For example, the small protein bovine pancreatic trypsin inhibitor (BPTI) (58 amino acid residues) consists of 568 atoms (hydrogen atoms bound to carbons excluded), and its dimensions are 28 Å by 28 Å by 40 Å. If we approximate it by a sphere of radius 16 Å and put it into a cube with a minimum solvent layer thickness of 6 Å to the walls, about 2300 water molecules (6900 atoms) are needed to fill the non-protein part of the cube of volume  $[2 \times (16 + 6)]^3 \text{ Å}^3$ . To fill the non-protein part of a truncated octahedron of volume  $\frac{1}{2}[4 \times (16 + 6)/\sqrt{3}]^3 \text{ Å}^3$  takes about 1600 water molecules (4800 atoms). This means a reduction of the system size by at least one quarter of the number of atoms.

### 3.3.3. Extended Wall Region Boundary Condition

If a molecular system is irregular, i.e. not nearly translationally periodic, periodic boundary conditions cannot be applied. In this case the distorting effect of the vacuum outside the molecular system may be reduced by designating a layer of atoms of the system to be an extended wall region, in which the motion of the atoms is restrained in order to avoid the deforming influence of the nearby vacuum (Fig. 10c, d). The atoms in the extended wall region can be kept fixed<sup>[89]</sup> or harmonically restrained to stationary positions. Their motion may be coupled to a heat bath, e.g. by

applying stochastic dynamics<sup>[89]</sup> in order to account for exchange of energy with the surroundings. In any case the type of force applied to these atoms should be chosen such that their motion in the finite system resembles as closely as possible the true motion in an infinite system. The extended wall region forms a buffer between the fully unrestrained part of the system and the (unrealistic) vacuum surrounding it.

This extended wall region technique has been applied in the simulation of solids,<sup>[104]</sup> liquids<sup>[105, 106]</sup> and proteins.<sup>[107, 108]</sup> Although the technique yields considerable savings in computing time, it should be carefully analyzed how far the restraining of the motion of the atoms in the wall region will affect the motion of the freely simulated atoms in the system.

## 3.4. Different Types of Molecular Dynamics

When Newton's equations of motion (39) and (40) are integrated the total energy is conserved (adiabatic system) and if the volume is held constant the simulation will generate a microcanonical ensemble. For various reasons this is not very convenient and a variety of approaches has appeared in the literature to yield a type of dynamics in which temperature and pressure are independent variables rather than derived properties. When MD is performed in non-equilibrium situations in order to study irreversible processes, catalytic events or transport properties, the need to impress external constraints or restraints on the system is apparent. In such cases the temperature should be controlled as well in order to absorb the dissipative heat produced by the irreversible process. But also in equilibrium simulations the automatic control of temperature and pressure as independent variables is very convenient. Slow temperature drifts that are an unavoidable result of force truncation errors are corrected, while also rapid transitions to new desired conditions of temperature and pressure are more easily accomplished.

### 3.4.1. Constant-Temperature Molecular Dynamics

Several methods for performing MD at constant temperature have been proposed, ranging from ad-hoc rescaling of atomic velocities in order to adjust the temperature, to consistent formulation in terms of modified Lagrangian equations of motion that force the dynamics to follow the desired temperature constraint. Different types of methods can be distinguished.

1. *Constraint methods*, in which the current temperature  $T(t)$  at time point  $t$  is exactly equal to the desired reference temperature  $T_0$ . This can be achieved by rescaling the velocities at each MD time step<sup>[20]</sup> by a factor  $[T_0/T(t)]^{1/2}$ , where the temperature  $T(t)$  is defined in terms of the kinetic energy through equipartition [Eq. (43)]. The number of degrees of freedom in the system is  $N_{df}$ .

$$E_{kin}(t) = \sum_{i=1}^N \frac{1}{2} m_i v_i^2(t) = \frac{1}{2} N_{df} k_B T(t) \quad (43)$$

The same result can be obtained in a more elegant way by a modification of the equations of motion (39),<sup>[109–111]</sup> such

that the kinetic energy instead of the total energy becomes a constant of motion. This method has two disadvantages. First, numerical inaccuracy of the algorithm may produce a drift in temperature that is not stabilized, since the reference temperature  $T_0$  does not appear in the equations to be integrated. Second, although the system is modeled by a Hamiltonian, this Hamiltonian does not represent a physical system, for which fluctuations in the kinetic energy are characteristic. In practical applications which are aimed at realistically simulating a physical system, the use of a non-physical Hamiltonian is questionable, even though mathematically consistent equations of motion are obtained.

2. *Extended system methods*,<sup>[112]</sup> in which an extra degree of freedom  $s$  representing a heat bath is added to the atomic degrees of freedom of the molecular system. A kinetic and a potential energy term involving this extra degree of freedom are added to the Hamiltonian and the simulation is carried out for the  $N_{\text{df}} + 1$  degrees of freedom of this extended system. Due to the presence of a kinetic term  $\frac{1}{2}m_s(ds/dt)^2$  in the Hamiltonian, energy is flowing dynamically from the heat bath to the system and back, the speed being controlled by the inertia parameter  $m_s$ . A disadvantage of this type of second-order coupling to a heat bath is the occurrence of spurious energy oscillations, the period of which depends on the value of the adjustable coupling parameter  $m_s$ .

3. *Weak coupling methods*,<sup>[113]</sup> in which the atomic equations of motion are modified such that the net result on the system is a first-order relaxation of the temperature towards the reference value  $T_0$  [Eq. (44)].

$$dT(t)/dt = \tau_T^{-1} [T_0 - T(t)] \quad (44)$$

The kinetic energy can be changed by  $\Delta E_{\text{kin}}$  in a MD time step  $\Delta t$  by scaling all atomic velocities  $\vec{v}_i$  with a factor  $\lambda$ . Using (43) we get (45).

$$\Delta E_{\text{kin}} = (\lambda^2 - 1) \frac{1}{2} N_{\text{df}} k_B T(t) \quad (45)$$

If the heat capacity per degree of freedom of the system is denoted by  $c_V^{\text{df}}$  the change in energy [Eq. (45)] leads to a

$$\Delta T = [N_{\text{df}} c_V^{\text{df}}]^{-1} \Delta E_{\text{kin}} \quad (46)$$

change in temperature [Eq. (46)], which should be equal to  $\Delta T$  as determined by (44). Solving Equations (44)–(46) for  $\lambda$  we obtain equation (47).

$$\lambda = [1 + c_V^{\text{df}} (k_B/2)^{-1} \Delta t \tau_T^{-1} (T_0/T(t) - 1)]^{1/2} \quad (47)$$

The heat capacity per degree of freedom  $c_V^{\text{df}}$  may not be accurately known for the system. This has no consequence for the dynamics since the temperature relaxation time  $\tau_T$  is an adjustable parameter. This aperiodic coupling to a heat bath through a first-order process has the advantage over the extended system methods that the response to temperature changes is non-oscillatory and that it is rather easy to implement through a simple velocity scaling using Equation (47). The coupling can be chosen sufficiently weak (sufficiently large  $\tau_T$ ) to avoid disturbance of the system and sufficiently strong (small  $\tau_T$ ) to achieve the desired result.

4. *Stochastic methods*, in which the individual atomic velocities  $\vec{v}_i$  are changed stochastically. Andersen<sup>[114]</sup> proposed a Maxwellian re-thermalization procedure by stochastic collisions, in which the mean time between collisions plays the role of adjustable parameter determining the strength of the coupling to the heat bath. Heyes<sup>[115]</sup> has suggested a Monte Carlo type technique for selection of new velocities. Others<sup>[116]</sup> used the Langevin equation (41) to achieve the coupling to a heat bath, the strength of the coupling being determined by the value of the atomic friction coefficients  $\gamma_i$ .

### 3.4.2. Constant-Pressure Molecular Dynamics

For an isotropic system the pressure is a scalar defined by Equation (48), where  $V$  denotes the volume of the computational box and the virial  $\Xi$  is defined as in Equation (49).

$$P = 2/(3V) [E_{\text{kin}} - \Xi] \quad (48)$$

$$\Xi = -\frac{1}{2} \sum_{\text{pairs } (i,j)}^N \mathbf{r}_{ij} \cdot \mathbf{F}_{ij} \quad (49)$$

Here,  $\mathbf{r}_{ij} = \mathbf{r}_i - \mathbf{r}_j$  and  $\mathbf{F}_{ij}$  is the force on atom  $i$  due to atom  $j$ . For molecular systems, forces within a molecule may be omitted together with kinetic energy contributions of intramolecular degrees of freedom. A pressure change can be achieved by scaling of the volume of the box and by changing the virial through a scaling of interatomic distances.

The various methods for carrying out MD at constant pressure are based on the same principles as the constant temperature scheme with the role of the temperature played by the pressure and the role of the atomic velocities played by the atomic positions. The following methods can be distinguished.

1. *Constraint methods*,<sup>[117]</sup> in which the equations of motion are modified such that the pressure instead of the volume becomes a constant of motion. This type of method has the same two disadvantages as its constant temperature counterpart.

2. *Extended system methods*,<sup>[114, 118–121]</sup> in which an extra degree of freedom  $V$ , the volume of the box, is added to the atomic degrees of freedom of the system. A kinetic energy term,  $\frac{1}{2}m_v(dV/dt)^2$  and a potential energy term,  $PV$ , involving the extra degree of freedom are added to the Hamiltonian, and the simulation is carried out for the  $N_{\text{df}} + 1$  degrees of freedom of the extended system. The rate of volume change is governed by the inertia parameter  $m_v$ . A disadvantage of this second-order coupling method is the occurrence of spurious volume oscillations, the period of which depends on the size of the adjustable parameter  $m_v$ .

3. *Weak coupling methods*,<sup>[113]</sup> in which the atomic equations of motion are modified such that the net result on the system is a first-order relaxation of the pressure towards a reference value  $P_0$  [Eq. (50)].

$$dP(t)/dt = \tau_P^{-1} [P_0 - P(t)] \quad (50)$$

Scaling the atomic coordinates  $\mathbf{r}_i$  and the edges of the computational box by a factor  $\mu$  leads to a volume change [Eq. (51)].

$$\Delta V = (\mu^3 - 1)V \quad (51)$$

The pressure change  $\Delta P$  due to this change in volume is given by Equation (52), where the isothermal compressibility of

$$\Delta P = -(\beta_T V)^{-1} \Delta V \quad (52)$$

the system is denoted by  $\beta_T$ . Solving Equations (50)–(52) for  $\mu$  we get Equation (53).

$$\mu = [1 - \beta_T \Delta t \tau_p^{-1} (P_0 - P(t))]^{1/3} \quad (53)$$

Since the pressure relaxation time  $\tau_p$  is an adjustable parameter, an accurate value for the compressibility of the system is not required. This first-order pressure coupling method has advantages corresponding to those of the weak temperature coupling method [cf. Eqs. (44)–(47)]. The expressions (48)–(53) can easily be modified to apply to a general anisotropic triclinic system. Virial, kinetic energy, pressure, and the scaling factor  $\mu$  will become cartesian tensors, and the volume  $V$  becomes the determinant of the matrix formed by the vectors  $\mathbf{a}$ ,  $\mathbf{b}$  and  $\mathbf{c}$  representing the edges of the computational box.

4. *Stochastic methods* for constant-pressure MD have not yet been proposed. They would involve random changes of the box volume.

### 3.5. Algorithms for Molecular and Stochastic Dynamics

#### 3.5.1. Integration Schemes for Molecular Dynamics

Newton's equation of motion (39), a second-order differential equation, can be written as two first-order differential equations (54) and (55) for the particle positions  $\mathbf{r}_i(t)$  and

$$d\mathbf{v}_i(t)/dt = m_i^{-1} \mathbf{F}_i(\{\mathbf{r}_i(t)\}) \quad (54)$$

$$d\mathbf{r}_i(t)/dt = \mathbf{v}_i(t) \quad (55)$$

velocities  $\mathbf{v}_i(t)$  respectively. The forces  $\mathbf{F}_i$  are obtained from the potential energy function through Equation (40); they therefore depend on the configuration  $\{\mathbf{r}_i(t)\}$  of the system. A simple algorithm for integration of Equations (54) and (55) in small time steps  $\Delta t$  is obtained as follows.

Taylor expansion of  $\mathbf{v}_i(t)$  at time point  $t = t_n$  yields the expression (56) when Equation (54) is used.

$$\begin{aligned} \mathbf{v}_i(t_n + \Delta t/2) &= \mathbf{v}_i(t_n) + d\mathbf{v}_i/dt|_{t_n} \Delta t/2 \\ &\quad + d^2\mathbf{v}_i(t)/dt^2|_{t_n} (\Delta t/2)^2/2! + O(\Delta t^3) \\ \mathbf{v}_i(t_n - \Delta t/2) &= \mathbf{v}_i(t_n) - d\mathbf{v}_i/dt|_{t_n} \Delta t/2 \\ &\quad + d^2\mathbf{v}_i(t)/dt^2|_{t_n} (\Delta t/2)^2/2! + O(\Delta t^3) \\ \hline \mathbf{v}_i(t_n + \Delta t/2) &= \mathbf{v}_i(t_n - \Delta t/2) + m_i^{-1} \mathbf{F}_i(\{\mathbf{r}_i(t_n)\}) \Delta t + O(\Delta t^3) \end{aligned} \quad (56)$$

Using the same procedure for Taylor expansions of  $\mathbf{r}_i(t)$  at time point  $t = t_n + \Delta t/2$  and using Equation (55) we obtain Equation (57).

$$\mathbf{r}_i(t_n + \Delta t) = \mathbf{r}_i(t_n) + \mathbf{v}_i(t_n + \Delta t/2) \Delta t + O(\Delta t^3). \quad (57)$$

Equations (56)–(57) form the so-called leap-frog scheme. Its name is illustrated in Figure 11. It is one of the most accurate, stable, and yet simple and efficient algorithms available for molecular dynamics of fluid-like systems. Our

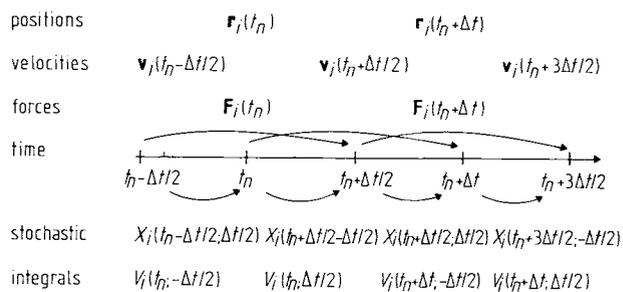


Fig. 11. The leap-frog scheme for integration of Newton's equations of motion and Langevin's equations of motion.

preference for using the leap-frog scheme instead of other algorithms such as those of *Runge-Kutta*,<sup>[122]</sup> *Gear*,<sup>[123]</sup> *Verlet*<sup>[124]</sup> or *Beeman*<sup>[125]</sup> is a result of performance evaluations<sup>[28, 69, 126, 127]</sup> and of the following considerations.

1. *Number of force evaluations per time step.* By far the most expensive part of a MD simulation is the force calculation [cf. Eq. (40)], which therefore should not be carried out more than once per integration time step  $\Delta t$ . This rules out algorithms of the Runge-Kutta type.

2. *Order of the algorithm.* It can be shown<sup>[69, 127]</sup> that the leap-frog, Verlet and Beeman algorithms generate exactly the same trajectory, and are of 3rd-order accuracy in the time step  $\Delta t$ . The application of higher-order, more accurate algorithms like those of *Gear*,<sup>[123]</sup> which involve the use of higher-order derivatives of the function to be integrated, is of no use as long as this function is a non-harmonic, noisy one. For example, in simulations of polar liquids or macromolecular solutions noise due to the cut-off applied to the long-range Coulomb forces prohibits an increase in accuracy beyond 3rd order.<sup>[126]</sup> However, if the highest-frequency motions in the molecular system are highly harmonic, as in solids, the higher-order function derivative will possess predictive power, which may lead to improved accuracy by applying higher-order algorithms. Algorithms that are of lower than third order are not efficient in the MD of molecular systems.<sup>[127]</sup> This is because molecular potential energy functions  $V$  generally have positive second derivatives, requiring third- or higher-order algorithms.

3. *Representation of the algorithm.* The much used Verlet algorithm can be obtained from the leap-frog scheme [(56)–(57)] by eliminating the velocities  $\mathbf{v}_i(t_n + \Delta t/2)$  and  $\mathbf{v}_i(t_n - \Delta t/2)$  from Equations (56) and (57) and replacement of  $t_n$  by  $t_n - \Delta t$  in Equation (57).

$$\begin{aligned} \mathbf{r}_i(t_n + \Delta t) &= 2\mathbf{r}_i(t_n) - \mathbf{r}_i(t_n - \Delta t) \\ &\quad + m_i^{-1} \mathbf{F}_i(\{\mathbf{r}_i(t_n)\}) (\Delta t)^2 + O(\Delta t^4) \end{aligned} \quad (58)$$

The atomic velocity does not occur explicitly in this algorithm, which makes a coupling of the system to a heat bath by velocity scaling as described in Section 3.4 impossible.

4. *Memory and computational requirements.* In some studies the Beeman algorithm [Eqs. (59) and (60)] is advocated

$$\mathbf{r}_i(t_n + \Delta t) = \mathbf{r}_i(t_n) + \mathbf{v}_i(t_n)\Delta t + m_i^{-1} [4\mathbf{F}_i(t_n) - \mathbf{F}_i(t_n - \Delta t)](\Delta t)^2/6 \quad (59)$$

$$\mathbf{v}_i(t_n + \Delta t) = \mathbf{v}_i(t_n) + m_i^{-1} [2\mathbf{F}_i(t_n + \Delta t) + 5\mathbf{F}_i(t_n) - \mathbf{F}_i(t_n - \Delta t)]\Delta t/6 \quad (60)$$

ed.<sup>[44]</sup> Although Equations (59) and (60) are much more complicated than Equation (58), the latter follows directly from the former: replace  $t_n$  by  $t_n - \Delta t$  in Equations (59) and (60), multiply the latter by  $\Delta t$  and subtract the former from it, and add the resulting equation to Equation (59). Reshuffling of terms then gives Equation (58). Since the Beeman, Verlet and leap-frog algorithms generate identical trajectories, we prefer to use the latter because of its minimal computer memory storage and computational requirements.

### 3.5.2. Integration Schemes for Stochastic Dynamics

Langevin's equation of motion (41) differs from Newton's equation (54) by the occurrence of a stochastic force  $\mathbf{R}_i(t)$  and a frictional force  $m_i\gamma_i\mathbf{v}_i(t)$  [Eq. (61)].

$$d\mathbf{v}_i(t)/dt = m_i^{-1}\mathbf{F}_i(\{\mathbf{r}_i(t)\}) + m_i^{-1}\mathbf{R}_i(t) - \gamma_i\mathbf{v}_i(t) \quad (61)$$

The solution of this equation around  $t = t_n$  is formulated in Equation (62).

$$\mathbf{v}_i(t) = \mathbf{v}_i(t_n)e^{-\gamma_i(t-t_n)} + m_i^{-1}e^{-\gamma_i(t-t_n)}\int_{t_n}^te^{-\gamma_i(t_n-t')}[\mathbf{F}_i(t') + \mathbf{R}_i(t')]dt' \quad (62)$$

Since the stochastic properties of  $\mathbf{R}_i(t')$  are given, the integral over  $\mathbf{R}_i(t')$  can be obtained directly. The integral over the systematic force  $\mathbf{F}_i(t')$  is, as in the previous section, obtained by expanding  $\mathbf{F}_i(t')$  in a Taylor series around  $t = t_n$  and omitting all terms beyond third order in  $\Delta t$  in the positions, beyond second order in the velocities, and beyond second order in the forces. The SD equivalent of the leap-frog velocity formula (56) then becomes:<sup>[128]</sup>

$$\mathbf{v}_i(t_n + \Delta t/2) = \mathbf{v}_i(t_n - \Delta t/2)e^{-\gamma_i\Delta t} + m_i^{-1}\mathbf{F}_i(t_n)[1 - e^{-\gamma_i\Delta t}]/(\gamma_i\Delta t) + \mathbf{V}_i(t_n; \Delta t/2) - e^{-\gamma_i\Delta t}\mathbf{V}_i(t_n; -\Delta t/2) \quad (63)$$

with

$$\mathbf{V}_i(t_n; \Delta t/2) \equiv m_i^{-1}e^{-\gamma_i\Delta t/2}\int_{t_n}^{t_n+\Delta t/2}e^{-\gamma_i(t_n-t')}\mathbf{R}_i(t')dt' \quad (64)$$

When  $\gamma_i$  goes to zero, Equation (63) reduces to Equation (56). The SD equivalent of the leap-frog position formula (57) is obtained by integrating Equation (55) using the expression (62) for the velocity<sup>[128]</sup> [Eqs. (65), (66)].

$$\mathbf{r}_i(t_n + \Delta t) = \mathbf{r}_i(t_n) + \mathbf{v}(t_n + \Delta t/2)\Delta t[e^{+\gamma_i\Delta t/2} - e^{-\gamma_i\Delta t/2}]/(\gamma_i\Delta t) + \mathbf{X}_i(t_n + \Delta t/2; \Delta t/2) - \mathbf{X}_i(t_n + \Delta t/2; -\Delta t/2) \quad (65)$$

$$\mathbf{X}_i(t_n; \Delta t/2) \equiv (m_i\gamma_i)^{-1}\int_{t_n}^{t_n+\Delta t/2}[1 - e^{-\gamma_i(t_n+\Delta t/2-t')}] \mathbf{R}_i(t')dt' \quad (66)$$

When  $\gamma_i$  goes to zero, Equation (65) reduces to Equation (57).

When using the SD leap-frog algorithm (63–66) it must be noted that  $\mathbf{V}_i(t_n; -\Delta t/2)$  is correlated with  $\mathbf{X}_i(t_n - \Delta t/2; \Delta t/2)$ , since they are different integrals of  $\mathbf{R}_i(t)$  over the time interval  $(t_n - \Delta t/2; t_n)$  (see Fig. 8). The same observation holds for  $\mathbf{X}_i(t_n + \Delta t/2; -\Delta t/2)$  and  $\mathbf{V}_i(t_n; \Delta t/2)$  [cf. Eqs. (65) and (63)]. These are different integrals of  $\mathbf{R}_i(t)$  over the time interval  $(t_n, t_n + \Delta t/2)$ . This means that these correlated quantities must be sampled in a correlated manner,<sup>[128]</sup> which makes the SD leap-frog algorithm more complicated than the MD one.

Integration schemes for the generalized Langevin equation involving time-dependent friction coefficients  $\gamma_i$  can be found in Refs. [129, 130].

### 3.5.3. Application of Constraints in Molecular Dynamics

In Section 3.4 methods to exactly constrain the temperature and pressure of a molecular system were briefly mentioned. Here, methods to constrain molecular bond lengths or bond angles in dynamic simulations will be discussed. They are used to save computing time. The length of the time step  $\Delta t$  in a MD or SD simulation is limited by highest frequency ( $\nu_{\max}$ ) motions occurring in the system [Eq. (67)].

$$\Delta t \ll \nu_{\max}^{-1}. \quad (67)$$

By freezing the generally uninteresting high-frequency internal vibrations, such as bond-length or possibly bond-angle vibrations,  $\nu_{\max}^{-1}$  is increased, which allows for a longer time step  $\Delta t$ . The application of constrained dynamics makes sense physically and computationally when

1. the frequencies of the frozen (constrained) degrees of freedom are (considerably) higher than those of the remaining ones, thereby allowing a (considerable) increase of  $\Delta t$ ,
2. the frozen degrees of freedom are only weakly coupled to the remaining ones, i.e. when the molecular motion is not significantly affected by application of constraints,
3. so-called metric tensor effects<sup>[131]</sup> play a minor role,
4. the algorithm by which the constraints are imposed on the molecular system does not require excessive mathematical or computational effort.

In molecular simulations typically a factor of 3 in computer time can be saved by application of constraints.<sup>[126]</sup> The application of constraints implies that the atoms move on a hypersurface in configuration space. While in the full Cartesian configuration space each configuration has equal weight in the partition function, the weighting on the hypersurface will in general depend on its form. The metric tensor defining the hypersurface in terms of Cartesian coordinates will determine these weights. Therefore, when simulating using constraints, metric tensor corrections must be taken into account. Their significance depends on the type of constraint. A discussion of metric tensor effects is given in Refs. [131, 132].

### 3.5.3.1. Classification of Methods

Several methods exist for integrating the equations of motion of a molecular system in the presence of constraints. They can be classified as follows:

1. *Formulation in terms of generalized coordinates.* In this case the constrained degrees of freedom (e.g. bond lengths, bond angles) are treated as fixed parameters, not as degrees of freedom. For example, a macromolecule may be modeled by only torsional angle degrees of freedom, as is often done in static modeling studies. Two cases can be distinguished.
  - a. When dealing with *rigid molecules*, center of mass coordinates and Euler angles can be used as degrees of freedom, leading to the Newton-Euler equations of rigid body motion. A variety of algorithms have been proposed for the integration of these equations.<sup>[19, 133–135]</sup>
  - b. For the more general case of *flexible molecules*, especially when the number of internal degrees of freedom becomes larger than a few, and when inertial terms are not neglected, it is a tedious task to write down explicitly the appropriate equations of motion in generalized coordinates.<sup>[136, 137]</sup> The reason is that the base vectors of the coordinate system become dependent on time, which has a number of unpleasant consequences.<sup>[136]</sup> So we think that the use of generalized coordinates, such as torsional angles, is highly unpractical for describing the dynamics of macromolecules.
2. *Formulation in terms of Cartesian coordinates.* Two methods are available for integrating the Cartesian equations of motion of flexible molecules subject to holonomic scleronomous constraints, that is, constraints that are only dependent on atomic coordinates, not on time.
  - a. *Matrix methods*<sup>[138]</sup> bear their name because they involve the (costly) inversion of a matrix of dimension equal to the number of constraints. Thus, these methods are not well suited for application to macromolecular systems.
  - b. *Iterative methods*, like the so-called SHAKE method,<sup>[138]</sup> are especially appropriate for macromolecules, since they treat the constraints in an iterative way.

### 3.5.3.2. Algorithms for Constraint Dynamics in Cartesian Coordinates

Molecular constraints have the form (68) for the case of

$$\sigma_k(\mathbf{r}_1, \dots, \mathbf{r}_N) = 0 \quad k = 1, \dots, N_c \quad (68)$$

$N_c$  constraints in a molecule consisting of  $N$  atoms. Bond-length and bond-angle constraints can be put in the form of distance constraints between atoms  $k_1$  and  $k_2$ , as in (69) where the constraint distance is given by  $d_{k_1 k_2}$ .

$$r_{k_1 k_2}^2 - d_{k_1 k_2}^2 = 0 \quad (69)$$

When applying constraints in MD or SD, the  $3N$  equations of motion (39) or (41) have to be integrated while satisfying the  $N_c$  constraints. This can be accomplished by applying Lagrange's method of undetermined multipliers.<sup>[122]</sup> A zero term (68) is added to the potential energy

function  $V$  in Equation (40), which yields (70) as the equation of motion.

$$m_i d^2 \mathbf{r}_i(t)/dt^2 = - \frac{\partial}{\partial \mathbf{r}_i} \left\{ V(\{\mathbf{r}_i(t)\}) + \sum_{k=1}^{N_c} \lambda_k(t) \sigma_k(\{\mathbf{r}_i(t)\}) \right\} \quad (70)$$

The time-dependent multipliers  $\lambda_k(t)$  are determined such that the constraints  $\sigma_k$  are satisfied. The physical interpretation of Equation (70) becomes clear by rewriting it in terms of forces [Eq. (71)]:

$$m_i d^2 \mathbf{r}_i(t)/dt^2 = \mathbf{F}_i(t) + \mathbf{G}_i(t) \quad (71)$$

The total unconstrained force  $\mathbf{F}_i(t)$  derived from the potential energy function is the first term in Equation (70), while the constraint force  $\mathbf{G}_i(t)$ , which compensates the components of  $\mathbf{F}_i(t)$  that act along the directions of the constraints, is the second term.

The leap-frog scheme [(56)–(57)] for integration of Equation (71) becomes Equations (72) and (73).

$$\mathbf{v}_i(t_n + \Delta t/2) = \mathbf{v}_i(t_n - \Delta t/2) + m_i^{-1} \{\mathbf{F}_i(t_n) + \mathbf{G}_i(t_n)\} \Delta t \quad (72)$$

$$\mathbf{r}_i(t_n + \Delta t) = \mathbf{r}_i(t_n) + \mathbf{v}_i(t_n + \Delta t/2) \Delta t \quad (73)$$

Separating contributions from  $\mathbf{F}_i(t_n)$  and  $\mathbf{G}_i(t_n)$  we find

$$\mathbf{r}_i(t_n + \Delta t) = \mathbf{r}'_i + \delta \mathbf{r}_i \quad (74)$$

with

$$\delta \mathbf{r}_i = m_i^{-1} \mathbf{G}_i(t_n) (\Delta t)^2 \quad (75)$$

where  $\mathbf{r}'_i$  are the positions after a MD or SD step disregarding all constraints, and  $\delta \mathbf{r}_i$  are the positional corrections to be made as a result of the constraints. Using the definition of  $\mathbf{G}_i$  we find

$$\delta \mathbf{r}_i = - m_i^{-1} (\Delta t)^2 \sum_{k=1}^{N_c} \lambda_k(t_n) \frac{\partial}{\partial \mathbf{r}_i} \sigma_k(\{\mathbf{r}_i(t_n)\}) \quad (76)$$

or when using the explicit form (69) for  $\sigma_k$ :

$$\delta \mathbf{r}_i = - 2 m_i^{-1} (\Delta t)^2 \sum_{\substack{k=1 \\ (k_1, k_2) = (i, j)}}^{N_c} \lambda_k(t_n) \mathbf{r}_{ij}(t_n) \quad (77)$$

where the summation extends only over constraints involving atom  $i$ . This implies that corrections due to the distance constraint between atoms  $i$  and  $j$  must be applied in the directions of the vector  $\mathbf{r}_{ij}$ . Corrections to  $\mathbf{r}'_i$  and  $\mathbf{r}'_j$  are in opposite directions and weighted by the inverse mass of atoms  $i$  and  $j$ , as is illustrated in Figure 12. Since the positions  $\mathbf{r}_i(t_n + \Delta t)$  must satisfy the constraints (68, 69) using (74) we have for each constraint:

$$[\mathbf{r}'_{k_1} + \delta \mathbf{r}_{k_1} - \mathbf{r}'_{k_2} - \delta \mathbf{r}_{k_2}]^2 = d_{k_1 k_2}^2 \quad (78)$$

These form a set of  $N_c$  quadratic equations from which the  $N_c$  Lagrangian multipliers  $\lambda_k$  can be determined. After linearization of Equation (78) by neglecting the terms quadratic

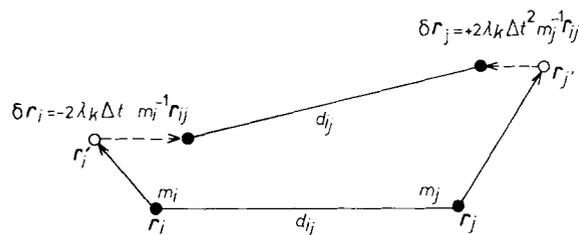


Fig. 12. Application of distance constraints. Atomic coordinate resetting using Lagrangian multipliers  $\lambda_k$ .

in  $\lambda_k$ , a set of linear equations is obtained. The above mentioned matrix methods solve these by a matrix inversion.<sup>[138]</sup> The iterative methods solve them by treating all constraints in succession, and iterating this procedure until all constraints are satisfied within a specific geometric tolerance.

The so-called SHAKE method, which is much used, is of the iterative type.<sup>[138]</sup> Its application in various algorithms is denoted by Equation (79). This means that the positions

$$\text{SHAKE } (r_i(t_n), r'_i(t_n + \Delta t), r_i(t_n + \Delta t)) \quad (79)$$

$r'_i(t_n + \Delta t)$  resulting from a non-constraint time step will be reset to give the constrained positions  $r_i(t_n + \Delta t)$ . When necessary the constraint forces can be determined from Equation (80) and the constrained velocities from Equation (81).

$$G_i(t_n) = m_i[r_i(t_n + \Delta t) - r'_i(t_n + \Delta t)]/(\Delta t)^2 \quad (80)$$

$$v_i(t_n + \Delta t/2) = [r_i(t_n + \Delta t) - r_i(t_n)]/\Delta t \quad (81)$$

### 3.5.3.3. Effect of the Application of Constraints

The effect of constraining bond lengths and bond angles in molecular systems has been evaluated.<sup>[126, 139]</sup> It turns out that the application of bond-length constraints saves about a factor of 2 in computing effort when bonds to hydrogen atoms are constrained and about a factor of 3 when all covalent bonds are constrained. No evidence was found for a distortion of the physical properties by the rigidity of the bonds.<sup>[139]</sup> When the bond angles undergo limited variation, which is generally true for molecular systems, metric tensor corrections play an insignificant role when bond-length constraints are applied.<sup>[132]</sup>

The use of bond-angle constraints is not allowed, since it considerably affects the molecular motion. Macromolecular flexibility and entropy are halved, and the number of torsional angle transitions is dramatically reduced.<sup>[139]</sup> Moreover, it has been shown that in case of application of bond angle constraints metric tensor corrections, which are nearly impossible to calculate for all but the smallest flexible molecules, are of significant size and may not be ignored.<sup>[132]</sup> This means that MD or SD of molecular systems should not be performed in torsional space while treating bond lengths and angles as fixed quantities.

### 3.5.3.4. Disadvantages of the Application of Constraints

The advantage of the application of bond-length constraints is clear: at the expense of about 10% extra comput-

ing time, longer, time steps  $\Delta t$  can be taken. When simulating macromolecular systems without constraints, a value of  $\Delta t = 0.5$  fs is appropriate, whereas with bonds to hydrogen atoms constrained  $\Delta t = 1.0$  fs, and with all bonds constrained  $\Delta t = 2.0$  fs is appropriate. So, a factor of 2 to 4 in computing effort is saved.

Yet, the application of constraints also has its disadvantages.

1. *Convergence problems for large planar groups.* In practice, procedures like the SHAKE method sometimes fail to converge to a molecular configuration satisfying all constraints. This is often due to the fact that the constraint forces act along the bond directions in the previous MD step (Fig. 12). For a planar group of atoms the constraint forces will act along vectors in the plane. So, when the other forces, e.g. due to charge repulsion, act orthogonally to the plane, it is nearly impossible for the constraint forces to counteract these, since they act orthogonally to each other. A solution to this problem has been proposed in Ref. [140], where virtual atoms lying outside the plane of the planar group are used to obtain constraint forces with sizeable components orthogonal to the plane of the group of real atoms. The implementation of this virtual atom technique would require definition of virtual atoms for each possible type of planar group. This makes this technique not very attractive for application in macromolecules.

2. *Free energy of creation or annihilation of atoms.* When free energy differences are computed using the coupling parameter approach, the application of constraints is a complicating factor. The Hamiltonian, or the potential energy function, is made a function of a coupling parameter  $\lambda$ , which is smoothly changed in the course of a simulation in order to obtain the work done by the system over a reversible change of  $\lambda$ . The problem now is that a constrained bond length cannot be removed from the system as a smooth function of  $\lambda$ . When computing the free energy of breaking a bond one faces the fact that the removal of a constraint is a discontinuous process. Even when only the length of a bond is changed (not removed) as a function of  $\lambda$ , computation of the free energy change due to the work done by the constraint forces is not simple, as is shown in Ref. [141].

3. *Computational bottleneck for parallelization of algorithms.* The constraint forces are generally computed after the other forces have been calculated and an unconstrained integration time step is performed. This makes the application of constraints a computational bottleneck: all other computations have to wait for its termination. Moreover, the most efficient methods for application of constraints, such as the SHAKE method, are of an iterative nature which makes them unsuitable for parallelization on a computer.

4. *Physical model of frozen bonds.* Although no significant effects due to freezing of bonds have been observed for molecular systems in equilibrium, it remains to be seen whether in non-equilibrium situations the finite flexibility of bonds may play a role in the dynamics, e.g. as an energy reservoir.

### 3.5.4. Multiple Time Step Algorithms

An alternative to the application of bond length constraints is the use of a multiple time step (MTS) integration

algorithm.<sup>[6, 142]</sup> The length of the integration time step is limited by the oscillation or relaxation time of the forces [see Eq. (67)]. In a molecular system three frequency ranges can be distinguished, as shown in Table 5: high-frequency bond-stretching forces  $F^{hf}$ , low-frequency long-range Coulomb forces  $F^{lf}$ , and the remaining intermediate frequency forces  $F^{if}$ . The contribution of the different forces to the atomic trajectories may be integrated using different time steps. An example is the twin range method discussed in Section 3.1.3.4, in which the long-range Coulomb force is kept constant during  $k$  time steps  $\Delta t$ , where  $k$  lies in the range 5–100.

Table 5. Various relaxation times in macromolecular systems, and force components to be used in multiple time step algorithms.

| Type of force  | Approximate oscillation or relaxation time (fs) | Force                                     |                       |                             |
|--|---|---|-----------------------|-----------------------------|
|  |   | Application at step $n' = 0, 1, 2, \dots$ | Update at step        | Size of force to be applied |
| 1. high frequency: $F^{hf}$<br>covalent bond stretching  | 10  | $t_n + n'\Delta t'$                       | $t_n + n'\Delta t'$   | $F_i^{hf}$                  |
| 2. intermediate frequency: $F^{if}$<br>bond-angle bending<br>dihedral angle torsion<br>van der Waals<br>shortrange Coulomb | 40  | $t_n + n'm\Delta t'$                      | $t_n + n'm\Delta t'$  | $mF_i^{if}$                 |
| 3. low frequency: $F^{lf}$<br>long-range Coulomb   |   | 1000                                      | $t_n + n'mk\Delta t'$ | $t_n + n'mk\Delta t'$       |

A multiple time step scheme could also be applied to the bond-stretching forces.<sup>[143]</sup> The conventional MD or SD time step  $\Delta t$  is subdivided into  $m$  ( $m = \text{odd}$ ) smaller substeps  $\Delta t'$ , so  $\Delta t = m\Delta t'$ . The covalent bond forces  $F^{hf}$  are evaluated at each time step  $t_n + n'\Delta t'$ , with  $n' = 0, 1, 2, \dots$ . When  $n'$  is a multiple of  $m$ , the other forces  $F^{if} + F^{lf}$  are applied, but multiplied by a factor  $m$  in order to compensate for their omission at the substeps. Application of this MTS algorithm to macromolecules shows that for  $m = 3$  or  $5$  the integration is as accurately performed as when bond-length constraints are used. Since the bond frequencies are about four times the other ones, a value of  $m = 4$  would be expected to be large enough to ensure proper integration of the bond forces.

We note that the physical quantities one is interested in should only be evaluated at the conventional time steps  $\Delta t$ , not at the substeps  $\Delta t'$ . Since the bond stretching forces can be rapidly computed, application of this MTS algorithm is as efficient as the use of the SHAKE method, but it avoids the disadvantages of the latter method which were discussed in the previous section.

### 3.5.5. Searching Neighbors

The bulk (approximately 90%) of the computer time required by a MD or SD simulation is used for calculating the nonbonded interactions, that is, for finding the nearest neighbor atoms and subsequently evaluating the van der Waals and Coulomb interaction terms for the atom pairs obtained. Various schemes for performing this task as efficiently as possible have been proposed.<sup>[6, 144]</sup>

1. *Neighbor list techniques.* Once the neighbors have been found, either by scanning of all possible atom pairs (operation proportional to  $N^2$ ), or by using grid-search techniques (proportional to  $N$ ), the pairs are stored in a neighbor list which is only updated every so many steps, typically every 5 to 100 steps. At each time step the neighbor list is used to calculate the interactions (operation proportional to  $N$ ).
2. *Grid search techniques.* The computational box is filled with a grid or mesh, and for each grid cell it is determined which atoms lie in it. This operation is proportional to  $N$ . The nearest neighbors of an atom can easily be found in the grid cells surrounding the grid cell containing the atom.

An evaluation of the different schemes can be found in Ref. [144]. For small systems ( $N \lesssim 1000$ ) the use of neighbor list techniques will reduce finding the neighbors to a small part of the calculation. For large systems ( $N \gtrsim 1000$ ) the application of grid search techniques will become advantageous.

## 3.6. Initial Conditions, Equilibration, and Analysis of Simulations

A simulation starts with initial atomic positions and velocities. The results should be independent of these. The initial configuration for a molecular system can be obtained from different sources: X-ray structure, model building, distance geometry calculation, random search techniques, etc. The initial velocities are either taken from a Maxwellian distribution or chosen to be zero, in which case strain in the molecule that is converted into kinetic energy may generate non-zero velocities.

The equilibration period that is required will depend on the relaxation time of the property one is interested in. Some properties, such as the kinetic energy, require short (picoseconds) equilibration times, whereas others, such as dielectric properties, may require longer times of the order of tens of picoseconds. During a simulation a number of quantities, such as the potential and kinetic energy, or the diffusion from the initial structure, are generally monitored to obtain a picture of the stability of the simulation.

The results are generally analyzed by taking time averages or averages over simulations with different initial conditions of the quantities of interest. Fluctuations and correlation functions may be calculated to analyze the mobility and dynamic behavior of the system.<sup>[6]</sup>

## 4. Application of Simulations

### 4.1. Understanding in Terms of Atomic Properties

The obvious utility of computer simulation is the possibility of analyzing molecular processes at the atomic level. Many examples can be found in the literature.<sup>[1–9]</sup> Here, we briefly mention two recent examples, the atomic interpretation of biochemical data on repressor-DNA operator binding,<sup>[26]</sup> and the atomic interpretation of biophysical measurements on membrane properties.<sup>[37]</sup>

#### 4.1.1. Interpretation of Biochemical Data on Repressor-Operator Binding by MD Computer Simulation

In Ref. [26] a 125-ps MD simulation of a Lac repressor headpiece (51 amino acid residues) complexed with its DNA operator (14 base pairs) in aqueous solution was reported. The starting structure for the simulation was obtained from model building and energy minimization based on 169 proton-proton distances for the headpiece and 24 headpiece-DNA proton-proton distances which were available from 2D-NMR measurements.<sup>[145]</sup> When analyzing the repressor-operator contacts, no evidence was found that a so-called "direct readout" mechanism<sup>[146]</sup> for recognition is based on direct repressor side-chain-base hydrogen bonds, which is the common view on recognition. The simulation suggested that direct readout occurs rather through non-polar contacts and water-mediated hydrogen bonds. The repressor-operator contacts as observed in the simulation are compatible with the available biochemical data on base-pair or amino acid residue substitution,<sup>[147, 148]</sup> but do not support the usual interpretation in terms of side-chain-base hydrogen bonds. This illustrates the usefulness of computer simulation studies to obtain an interpretation of biochemical data at the atomic level.

#### 4.1.2. Interpretation of Biophysical Data on Membrane Properties by MD Computer Simulation

In Ref. [37] a 180-ps MD simulation of a bilayered system of 52 sodium ions, 52 decanoate, 76 decanol and 526 water molecules was reported. A detailed analysis of the lipid-water interface was made. The charged (decanoate) head groups lie well within the water layer, but the alcoholic groups are situated more on the lipid side of the interface, which appears to be very diffuse and extends over almost 10 Å. Unexpectedly, the overlap of sodium ion and carboxylic acid distributions suggests a charge compensation rather than an electric double layer as pictured in most textbooks. Water molecule orientation is such that the remaining ion charge distribution is compensated.

Lateral diffusion constants of the lipid molecules as measured in the simulation ( $3 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$ ) compare well with experiment ( $2 \times 10^{-6} \text{ cm}^2 \text{ s}^{-1}$ ) as measured using nitroxide spin labels.<sup>[149]</sup> Analysis shows that the hydrodynamic interaction of the head groups with the aqueous layer determines the diffusion constant of the lipid molecules, rather than interactions within the lipid layer. This illustrates the usefulness of computer simulation studies to obtain an interpretation of biophysical data at the atomic level.

#### 4.2. Determination of Spatial Molecular Structure on the Basis of 2D-NMR, X-Ray or Neutron Diffraction Data

During the last few years computer simulation has become a standard tool in the determination of spatial molecular structure on the basis of X-ray or neutron diffraction or 2D-NMR data.<sup>[150]</sup> The goal of structure determination based on experimental NMR or diffraction data is to find a molecular structure that

1. satisfies the experimental data, such as a set of atom-atom distance constraints  $\{r_{ij}^0\}$  or torsional angle constraints  $\{\varphi_{ij}^0\}$  in the case of NMR, or a set of observed structure factor amplitudes  $F_{\text{obs}}(hkl)$  and, if available, phases  $\alpha(hkl)$  in the case of diffraction data, and
2. has a low energy in terms of a molecular potential energy function  $V_{\text{phys}}(\{r_i\})$  [Eq. (1)].

In order to optimize a structure simultaneously with respect to both criteria, the experimental information is cast into the form of a penalty function or restraining potential  $V_{\text{restr}}$ , the value of which increases the more an actual structure violates the experimental data. The most simple choice of a function taking into account maximum values for the interatomic distances  $\{r_{ij}^{\text{ub}}\}$  would be Equation (82), where the force constant is denoted by  $K_{\text{dr}}$ .

$$V_{\text{restr}} = V_{\text{dr}} \equiv \frac{1}{2} K_{\text{dr}} \sum_{\text{all constraints } (ij)} [\max(0, r_{ij} - r_{ij}^{\text{ub}})]^2 \quad (82)$$

The corresponding function which restrains the calculated structure factor amplitudes  $F_{\text{calc}}(hkl)$  to the observed ones is defined in (83).

$$V_{\text{restr}} = V_{\text{sf}} \equiv \frac{1}{2} K_{\text{sf}} \sum_{hkl} [F_{\text{calc}}(hkl) - F_{\text{obs}}(hkl)]^2 \quad (83)$$

The optimization problem is to find a molecular structure for which the energy function (84) attains the global

$$V_{\text{opt}} = V_{\text{phys}}(\{r_i\}) + V_{\text{restr}}(\{r_i\}) \quad (84)$$

minimum. As discussed in Section 3.2, MD simulation is a very powerful method to search configuration space for low energy configurations, due to its ability to surmount energy barriers of the order of  $k_B T$  per degree of freedom. Therefore, the application of MD computer simulation in spatial structure refinement of 2D-NMR or X-ray diffraction data has become widespread in recent years.

##### 4.2.1. Refinement of Structures Based on NMR Data

After introduction of the method of MD refinement<sup>[151]</sup> it has been applied to a variety of molecules using different refinement protocols.<sup>[150-154]</sup> The penalty function  $V_{\text{restr}}$  may be chosen in different ways.<sup>[154]</sup> The relative weight of the restraining term in Equation (84), that is the value of  $K_{\text{dr}}$ , may be chosen so large as to reduce the distance violations, at the expense of increasing the intramolecular energy  $V_{\text{phys}}$ . Too large a  $K_{\text{dr}}$  value will generally lead to a strained (unphysical) molecular structure.<sup>[28]</sup> The searching of configuration space may be performed at high temperature in order to more easily cross energy barriers.

The standard procedure is to start with a number of distance geometry (DG) type calculations to generate a collection of starting structures, which are subsequently refined by MD simulation. In Ref. [153] it was shown that the best DG structure in terms of distance violations is generally not the best structure after MD refinement. This is due to the crude energy function used in DG which has to be cast in the form of distances, and so favors geometries which are unfavored by more sophisticated energy functions used in simulations.

When the NMR data contain contributions from different molecular conformations it will be impossible to find *one* conformation satisfying the experimental data.<sup>[155]</sup> This observation led to the introduction of time-dependent constraints<sup>[156]</sup> which do not force the molecule to satisfy the distance constraints at each time point of the simulation, but only forces the distance constraints to be satisfied on average. This forms a better representation of the measured data.

#### 4.2.2. Refinement of Structure Based on Crystallographic X-Ray or Neutron Diffraction Data

Since its introduction the method of MD refinement of crystallographic data<sup>[157]</sup> has been applied mainly in protein crystallography. Again, different penalty functions (83) and refinement protocols are used. Generally speaking MD refinement saves a few months work over conventional refinement when applied to proteins. A survey of recent developments is given in Ref. [158].

As in the NMR case, the use of time-dependent structure factor restraints will be a much better representation of the experimental information, which is an average over time and over molecules. Preliminary results (*P. Gros*, private communication) indicate that lower *R* values and better searching of conformational space are obtained.

#### 4.3. Prediction of Structural Changes by MD Simulation

No reliable approaches are as yet available for the prediction of *de novo* (macro)molecular structures, and there is no real alternative to experimental structure determination. When structural information on homologous molecules is available, a molecular structure may be predicted starting from the homologous molecules and subsequently changing it to the required composition, either by model building on a graphic device or by more automatic simulation techniques. When additional experimental information from X-ray diffraction or NMR is available the chance of success is greatly increased. An example is the structure determination of the 279-residue protein thermitase, starting from the known structure of subtilisin (275 amino acids).<sup>[158, 159]</sup> These proteins show only 47% homology; residues had to be changed, deleted and inserted when changing subtilisin into thermitase. Subsequent MD refinement using crystallographic X-ray data on thermitase made numerous atoms move over more than 4 Å. Some parts showed a structural change of more than 8 Å.

If only one or a few amino acid side chains in a protein are changed (mutated), the structural change may be predicted by MD simulation without the extra help of X-ray data. In Figure 13 the conformation of the Met-222 → Phe mutant of subtilisin resulting from a MD simulation in which Met-222 was gradually changed into Phe-222<sup>[32]</sup> is compared with the conformation as obtained from an X-ray diffraction determination of this subtilisin mutant.<sup>[160]</sup> The predicted conformation is correct. However, this is not always true; especially when hydrogen bond networks, involving bound water molecules, have to be rearranged, the simulation period may be too short compared to the hydrogen bond network relaxation time.<sup>[32]</sup>

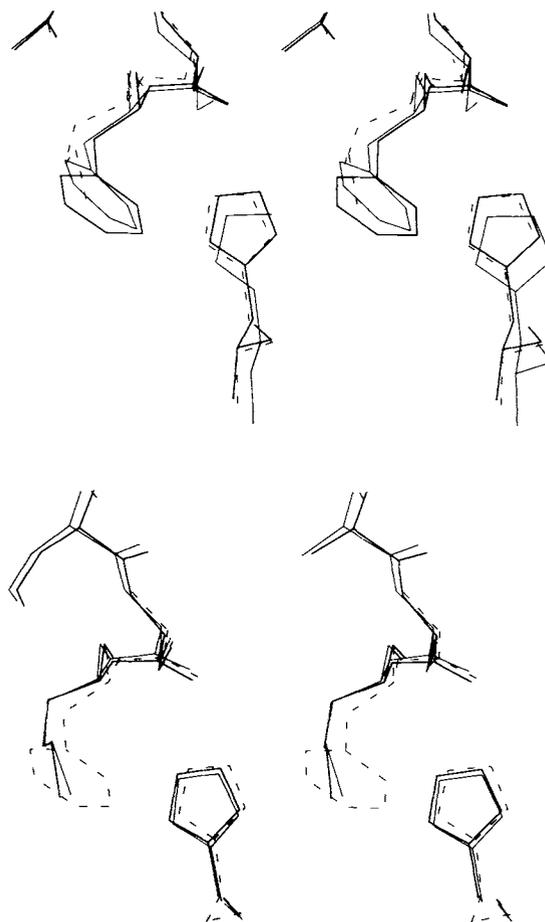


Fig. 13. Stereoview of the hydrophobic binding pocket of subtilisin. Upper part: forward 222 Met → Phe mutation; dashed line: 222 Met (native) starting structure; thick solid line: 222 Phe (mutant) MD predicted structure; thin solid line: 222 Phe (mutant) X-ray structure. Lower part: backwards 222 Phe → Met mutation; dashed line 222 Phe (MD) starting structure; thick solid line: 222 Met MD predicted structure (from 222 Phe mutant); thin solid line 222 Met (native) MD structure (from X-ray).

It is also possible by MD simulation to determine the conformation of a molecule as a function of the type of environment, viz. crystalline, nonpolar solution, aqueous solution, etc. Studies of this type have been performed for different molecules.<sup>[161–164]</sup> For the immuno suppressive drug cyclosporin A (CPA) the conformational differences found between its crystal conformation and its conformation in a chloroform solution were exactly reproduced by two MD simulations in the corresponding environments (Figure 16).<sup>[164]</sup>

This application illustrates the power of MD simulation when studying conformational properties of flexible molecules or mutants. However, when large conformational changes are to be expected the length (typically  $10^1$ – $10^2$  ps) of a MD simulation may be too short to bring these about in the limited time available to surmount possible energy barriers.

#### 4.4. Prediction of Free Energy Changes by MD Simulation

From an MD trajectory the statistical equilibrium averages can be obtained for any desired property of the molec-

ular system for which a value can be computed at each point of the trajectory. Examples of such properties are the potential and kinetic energy of relevant parts of the system, structural properties and fluctuations, electric fields, diffusion constants, etc. A number of thermodynamic properties can be derived from such averages. However, two important thermodynamic quantities, the entropy and the (Gibbs or Helmholtz) free energy, generally cannot be derived from a statistical average. They are global properties that depend on the extent of phase (or configuration) space accessible to the molecular system. Therefore, computation of the absolute free energy of a molecular system is virtually impossible. Yet, the most important chemical quantities like binding constants, association and dissociation constants, solubilities, adsorption coefficients, chemical potentials, etc., are directly related to the free energy. Over the past decade statistical mechanical procedures have evolved by means of which relative free energy differences may be obtained. For recent reviews of methodology and applications see Refs. [8, 165].

The most powerful method is the so-called thermodynamic cycle integration technique. The free energy difference between two states A and B of a system is determined from a MD simulation in which the potential energy function  $V(\{r_i\})$  [Eq. (1)] is slowly changed such that the system slowly converts from state A into state B. In principle the free enthalpy difference  $\Delta G_{BA} = G(B) - G(A)$  is determined as the work necessary to change the system from state A to state B via a reversible path. An example is given in Figure 14, where the free enthalpy of changing one water molecule (state A) in a cube with 216 water molecules into a methanol molecule (state B) is illustrated. The change is carried out over a period of 20 ps and yields  $\Delta G_{BA} = 6.1 \text{ kJ mol}^{-1}$ , as compared with an experimental value of  $5.2 \text{ kJ mol}^{-1}$ .

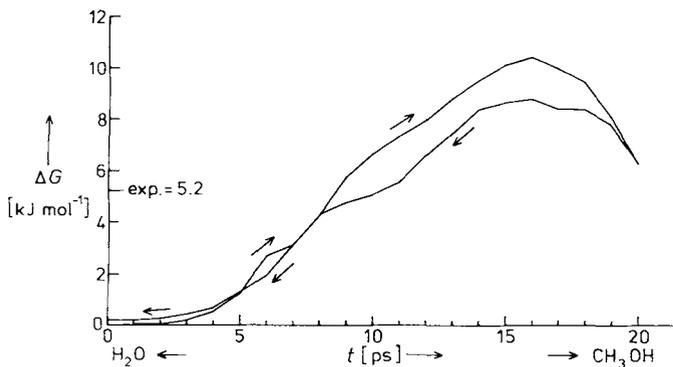


Fig. 14. Free enthalpy change  $\Delta G$  on changing  $\text{H}_2\text{O}$  into  $\text{CH}_3\text{OH}$  and back again.

The basis on which the thermodynamic cycle approach rests is the fact that the free enthalpy  $G$  is a thermodynamic state function. This means that as long as the system is changed in a reversible way the change in free enthalpy  $\Delta G$  will be independent of the path. Therefore, along a closed path or cycle one has  $\Delta G = 0$ . This result implies that there are two possibilities of obtaining  $\Delta G$  for a specific process; one may calculate it directly using the technique sketched above along a path corresponding to the process, or one may design a cycle of which the specific process is only a part, and

calculate the  $\Delta G$  of the remaining part of the cycle. The power of this thermodynamic cycle technique lies in the fact that on the computer also non-chemical processes such as the conversion of one type of atom into another type (H into  $\text{CH}_3$ , see Fig. 14) may be performed.

The method is outlined in Figure 15 for the process of binding of two different inhibitors, trimethoprim (TMP) and its triethyl analogue (TEP), to the enzyme dihydrofolate reductase (DHFR) from chicken liver<sup>[166, 167]</sup> in the presence of coenzyme (NADPH) and water. The appropriate thermodynamic cycle is given in Figure 15. The relative

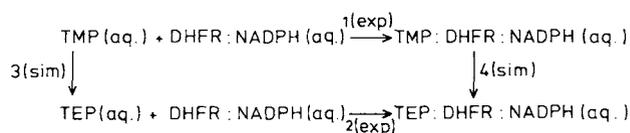
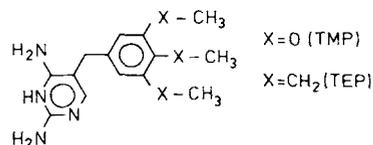


Fig. 15. Thermodynamic cycle for computation of the relative free enthalpy of binding of two inhibitors, trimethoprim (TMP) and its triethyl analogue (TEP), to the enzyme dihydrofolate reductase (DHFR) in the presence of coenzyme (NADPH) in aqueous solution. Complexation is denoted by the symbol “:”.

binding constant of the two inhibitors equals  $K_2/K_1 = \exp[-(\Delta G_2 - \Delta G_1)/RT]$ . However, simulation of processes 1 and 2 is virtually impossible, since it would involve the reversible removal of many solvent molecules from the active site of the enzyme to be substituted by the inhibitor. Since the processes in Figure 15 form a cycle, one has  $\Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3$ , and the processes 3 and 4 can easily be simulated since they involve the change of three oxygen atoms (TMP) into three  $\text{CH}_2$  groups (TEP). MD simulation of process 3 yields  $\Delta G_3 = -61 \pm 0.2 \text{ kJ mol}^{-1}$  and of process 4 yields  $\Delta G_4 = -65 \pm 10 \text{ kJ mol}^{-1}$ . So, the computed difference in free enthalpy of binding between TEP and TMP is  $+4 \pm 10 \text{ kJ mol}^{-1}$  to be compared with an experimental value of  $+7 \text{ kJ mol}^{-1}$ .

Although this result seems reasonable, the computed value may only be considered as an order-of-magnitude estimate. This is due to various assumptions and approximations that underlie this type of free energy calculation. Since these have been discussed elsewhere,<sup>[141]</sup> the most important ones will only be summarized here.

1. *Adequate sampling, or the relaxation time of the environment.* The change from state A to state B has to be carried out in a reversible way, which means that the time period of the change must be much longer than the relaxation time of the environment which has to adapt to the change. The rotational correlation time of a water molecule is about 2 ps and the dielectric relaxation time of water is about 8 ps. This means that a MD simulation of the change from A to B over 20 ps or more is long enough to obtain reasonably accurate  $\Delta G$  values (see Fig. 14). How-

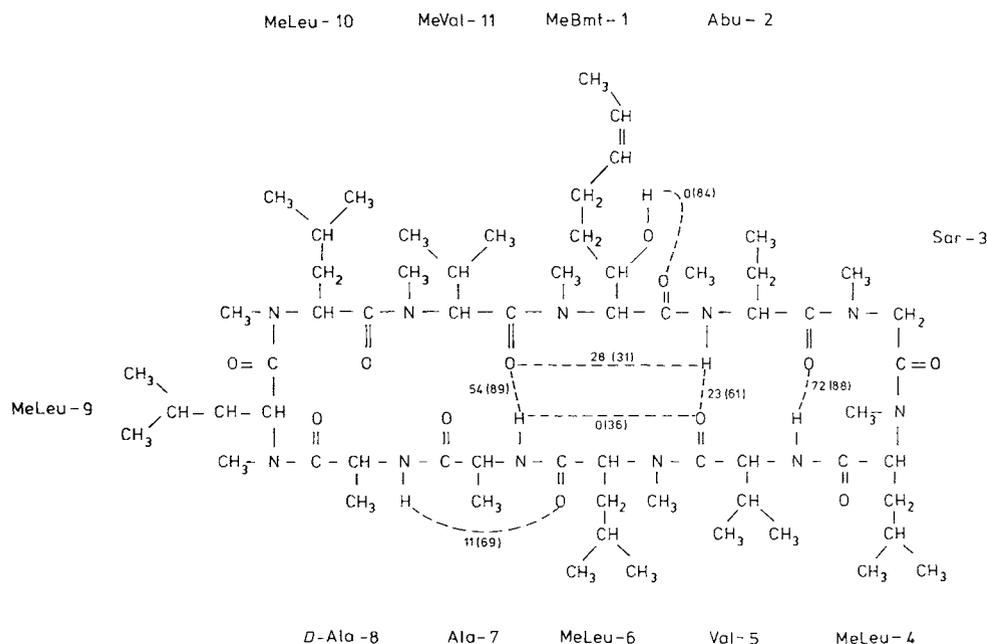


Fig. 16. Hydrogen bonds (dashed lines) occurring in two MD simulations of cyclosporin A. Percentage of occurrence is given, both for a MD simulation in aqueous solution and (in parentheses) for a MD simulation in vacuo.

ever, in process 4 (Fig. 15) the relaxation time of the (protein) environment of the inhibitor is much longer, which results in a large uncertainty of  $10 \text{ kJ mol}^{-1}$ .

2. *Effect of long-range Coulomb interactions.* Free energy computation involving the creation or annihilation of (full) atomic charges is very dependent on a proper treatment of the Coulomb interaction, which has an  $r^{-1}$  distance dependence. For example, when creating an ion in aqueous solution, the contribution of the hydration shell between  $9 \text{ \AA}$  and  $12 \text{ \AA}$  to the free energy of solvation is about  $40 \text{ kJ mol}^{-1}$ ,<sup>[41]</sup> which means that the cut-off radius must be very large to obtain accurate  $\Delta G$  values. When only dipolar changes play a role, as in the case of Figures 14 and 15, the interaction has an  $r^{-3}$  distance dependence, which considerably reduces the need to use a large cut-off radius.

3. *Sensitivity to force field parameters.* Not unexpectedly, the free energy of a molecular system is rather sensitive to the interaction function (1) that is used in the simulation. For example, in the relatively simple case of the free enthalpy of solvation of methanol in water (see Fig. 14), application of force field parameters of established force fields yields values ranging from 0 to  $14 \text{ kJ mol}^{-1}$ ,<sup>[28]</sup> a spread of  $2\text{--}3 kT$  around  $G_{\text{exp}} = 5.2 \text{ kJ mol}^{-1}$ .

The techniques for computation of (relative) free energy are in many applications not yet sufficiently accurate. However, they are still being improved and are in principle widely applicable in the study of molecular systems.

#### 4.5. Static versus Dynamic Representation of a Molecular System

The common approach to modeling a molecular system on a computer is a static one. For example, quantum calcu-

lations yield an equilibrium charge distribution, molecular mechanics calculations one or a few minimum energy conformations of a molecule; on a graphics device, molecules are studied in terms of fixed conformations. However, a molecular system at room temperature is by no means of static character. It should be described in terms of a multi-dimensional distribution function of all atomic coordinates and its development in time. An example of a dynamic equilibrium between different hydrogen bond patterns is given in Table 6 for crystalline cyclodextrin. The O61 atom is in-

Table 6. Dynamic equilibrium between various hydrogen bond patterns involving the O61 atom in crystalline  $\alpha$ -cyclodextrin [a]

| Donor [a] | Acceptor | %   | MD simulation                |                       | Neutron diffraction          |                       |
|-----------|----------|-----|------------------------------|-----------------------|------------------------------|-----------------------|
|           |          |     | distance<br>[ $\text{\AA}$ ] | angle<br>[ $^\circ$ ] | distance<br>[ $\text{\AA}$ ] | angle<br>[ $^\circ$ ] |
|           |          |     | D...A                        | D-H...A               | D...A                        | D-H...A               |
| O61-H61   | O63      | 79  | 2.79                         | 160                   | 2.81                         | 177                   |
|           | O65      | 2   | 3.04                         | 158                   |                              |                       |
|           | O56      | 1   | 2.99                         | 158                   |                              |                       |
|           | OW3      | 6   | 3.04                         | 146                   |                              |                       |
|           | OWA      | + 2 | 2.08                         | 156                   |                              |                       |
|           |          | 90  |                              |                       |                              |                       |
| O21-H21   | O63-H63  | 3   | 2.84                         | 161                   | 2.97                         | 173                   |
|           | OWA-HWA1 | 36  | 2.97                         | 156                   |                              |                       |
|           | OWA-HWA2 | 5   | 3.01                         | 154                   |                              |                       |
|           | OW3-HW31 | 8   | 3.07                         | 153                   |                              |                       |
|           | OW3-HW32 | 11  | 3.04                         | 153                   | 3.33                         | 107                   |
|           | OWB-HWB1 | + 2 | 2.94                         | 156                   |                              |                       |
|           |          |     | 79                           |                       |                              |                       |

[a] Abbreviations. D = donor, A = acceptor, W = water atoms [29].

involved in different, mutually exclusive hydrogen bonds. The ones that do exist more than about 20% of the MD simulation are also observed, as expected, in the neutron diffraction

experiment.<sup>[168]</sup> It is also observed that the lower the occurrence of a hydrogen bond is in the simulation, the more deformed its geometry is when described in terms of one static hydrogen bond, as is the case for the neutron diffraction data (see Table 6).

#### 4.6. The Role of a Solvent in Molecular Simulations

Many (macro)molecular modeling studies involve an isolated molecule without any solvent. This means that solvent effects are completely ignored. The molecular surface will be distorted by the vacuum boundary condition, and certainly no meaningful free energy estimates can be obtained. When a molecular complex is formed, solvent molecules may play a bridging role, as is observed in the repressor–operator complexes.<sup>[26, 169]</sup> Even intramolecular hydrogen bonding is affected by the competitive presence of water molecules surrounding a solute in aqueous solution. An example is given in Figure 16, where the occurrence of intramolecular hydrogen bonds is given for cyclosporin A, on the one hand simulated in vacuo and on the other in aqueous solution.<sup>[164]</sup> It is clear that omission of the solvent results in too high a percentage of intramolecular hydrogen bonds. Proper treatment of solvent effects is a necessary condition for a reliable simulation of molecular properties.

#### 4.7. Other Applications of Computer Simulation in Chemistry and Physics

The field of computer simulation of molecular systems has grown so rapidly that a review of all possible applications is hardly feasible. In this section we have chosen examples, mostly of our own work, to illustrate specific aspects of simulation studies. Many other applications may be found in a number of reviews and monographs on computer simulation in physics and chemistry.<sup>[1–9]</sup> We would like to mention a few recent studies which may lead the reader to areas of application that were not mentioned here. Computer simulation has been applied to the study of electrolytes,<sup>[170]</sup> ionic conductors,<sup>[171]</sup> ionic crystals,<sup>[172]</sup> ionic salts,<sup>[10]</sup> and, inter alia, to processes like adsorption,<sup>[173]</sup> sputtering<sup>[174]</sup> and melting.<sup>[175, 176]</sup>

### 5. Future Developments in Computer Simulation

#### 5.1. Quantum Simulations

Classical computer simulation implies a number of restrictions, as mentioned in Section 2.1.3, which are due to the assumption that quantum effects play a minor role. A proper description of low-temperature or light-atom (hydrogen) motion or chemical reactions requires a quantum mechanical treatment. Various quantum methods for application in the area of simulation have been developed and recent years have shown an increasing activity in the area of quantum simulations. Here, we briefly mention a few methods. Quantum effects can be incorporated in simulations in different ways.

- A. *Quantum corrections to the results of classical simulations* can be made. Expansion of the partition function in powers of  $\hbar$ , Planck's constant, leads to correction formulas for thermodynamic quantities like the free energy<sup>[177]</sup> or structural quantities like the radial distribution function  $g(r)$ .<sup>[178]</sup> Berens et al. propose a correction formula based on a harmonic approximation of the atomic motion.<sup>[179]</sup> Also for time-dependent equilibrium quantities, quantum correction formulas are known.<sup>[180]</sup>
- B. *Quantum mechanical treatment of a few degrees of freedom in an otherwise classical simulation* can be implemented in different ways.
  1. The *path-integral simulation method*<sup>[181–183]</sup> yields a quantum mechanical equilibrium distribution. Its name originates from its derivation by a discretization of the path-integral form of the density matrix. A recent application is the treatment of electron transfer reactions.<sup>[184]</sup>
  2. The *Gaussian wave packet method*<sup>[185–187]</sup> constituted an early attempt to solve the time-dependent Schrödinger equation by computer simulation. A problem inherent in this method is that of arriving at an adequate description of the interaction between a wave packet and its classical environment.
  3. In the *density functional dynamic method*<sup>[188]</sup> an energy functional, as used in density-functional theory, is added to the Lagrangian of the molecular system, and the Lagrangian equations of motion are subsequently integrated using MD techniques. Static and dynamic properties of crystalline silicon were obtained in terms of a self-consistent pseudo potential.<sup>[188]</sup>
  4. In the *adiabatic quantum molecular dynamics method*<sup>[189]</sup> the time-dependent Schrödinger equation for the excess electrons and Newton's equation of motion for the nuclei plus core electrons are integrated in the Born–Oppenheimer approximation. Electronic states and dynamic properties of dilute liquid K–KCl solutions were studied.<sup>[189]</sup>

The potential of the different methods has not yet been fully explored, which means that a best methodology has not as yet crystallized. The field is in a state of rapid development and holds much promise for a proper dynamic treatment of quantum degrees of freedom in the future.

#### 5.2. System Size and Time Scale of Simulations

The number of particles in a computer simulation typically lies in the range of  $10^2$ – $10^4$ , although simulations involving more than  $10^4$  atoms are nowadays feasible.<sup>[173]</sup> Relatively complex systems like a protein embedded in a membrane will also contain about  $10^5$  atoms. In practical applications limited system size is much less a concern than the finite time scale of a computer simulation. Due to large energy barriers in the potential energy surface, it may take a molecular system a very long time to cross these barriers and so to sample configuration space efficiently. Typical simulation periods are  $10^1$ – $10^2$  psec, which is much too short for a proper description of properties which show a much longer relaxation time. Therefore, possibilities to lengthen the time

scale of MD simulations are continuously investigated. Here we mention a few.

1. *Freezing of degrees of freedom.* The basic problem of removing degrees of freedom is that of defining an appropriate interaction function for the remaining degrees of freedom. Moreover, the degrees of freedom removed should play a minor role in the processes one is interested in. Application of bond-length constraints and the twin range method to handle long-range forces fall in this class. An alternative to freezing of degrees of freedom is the use of multiple time step integration methods.
2. *Stochastification of degrees of freedom.* Explicit treatment of degrees of freedom may be replaced by the application of a mean force plus a stochastic force which represent the average effect of the removed or stochastified degrees of freedom on the remaining explicitly treated ones. An example is the use of the Langevin equation to model the solvent or solid-state environment of a molecule.
3. *Scaling of system parameters.* Assume that we are interested in a quantity  $Q$  which has a relaxation time  $\tau$  that is longer than the time period that could be covered by a MD simulation. It may be possible to identify one or more system parameters  $p$  for which  $\tau$  is a (rapidly) changing function  $\tau(p)$  of  $p$ . In that case the physically correct value of  $p = p_{\text{phys}}$  can be changed to  $p_{\text{short}}$  such that  $\tau(p_{\text{short}})$  becomes shorter than the length of a MD simulation, which can then be carried out to obtain  $Q(p_{\text{short}})$ . Extrapolation of  $Q(p)$  for  $p$  from  $p_{\text{short}}$  to  $p_{\text{phys}}$  will yield an estimate  $Q(p_{\text{phys}})$ . The risk associated with the application of this technique is that by scaling of system parameters, the physical processes are changed such that a process at  $p_{\text{short}}$  may have nothing to do with one at  $p_{\text{phys}}$ . For example, when studying the relaxation of vibrational energy of an HCl molecule in an Ar lattice, the relaxation time may be shortened from the  $\mu\text{s}$  time scale to the ns time scale by reducing the HCl bond-stretching force constant. This will, however, certainly influence the balance between various relaxation mechanisms that are feasible for this process.
4. *Activated barrier crossing.* In the case of an activated process, methods exist to avoid a full simulation of the rare event of barrier crossing.<sup>[15]</sup> The procedure consists of three steps: 1) The location of the barrier must be determined. 2) The likelihood that the system will be at the top of the barrier is computed using umbrella sampling techniques. 3) The transition probability is computed by running MD trajectories from the top of the barrier. This technique has been applied to simulate ring flips in proteins.<sup>[16]</sup>
5. *Mass tensor dynamics.* In the classical partition function the integration over the momenta of the particles can be carried out separately from that over the coordinates, when no constraints are applied. The atomic masses do not appear in the configurational integral, which means that the equilibrium properties of a system will be independent of the masses in the system. The technique of mass tensor dynamics<sup>[190]</sup> exploits this freedom by choosing the atomic masses such that the high-frequency motions of the molecular system are slowed down, which allows a longer simulation time step to be taken.

None of the above-mentioned methods for lengthening the time scale of a simulation is really satisfactory. Most of the progress in this respect will probably result from the ever increasing power of computers. Yet, a combination of the different techniques mentioned above may yield a considerable reduction of the required computing power in suitable cases.

### 5.3. Accuracy of Molecular Model and Force Field

As was discussed in Section 3.1.1, there is no best molecular model or force field for all possible applications. The reliability of a particular force field will depend on the type of system and physical quantity it is applied to. Improvement of the quality and extension of the range of applicability is a continual concern in the area of computer simulation. Yet, the improvement of a force field is not a simple exercise for the following reasons.

1. When applying complicated force fields like (1), the exact *relationship between a force field parameter and a molecular property is often not known.* For example, how will the free enthalpy of solvation of methanol in water depend on the geometry and nonbonded interaction parameters?
2. *Force field parameters may be correlated.* For example, the actual barrier for a torsional rotation will depend on the combined effect of the dihedral angle potential energy term in (1) and the third neighbor nonbonded interaction between the first and the last atom defining the torsional angle.
3. There may be *conflicting requirements for improvement.* For example, the van der Waals radii of third neighbor atoms must not be too large in order that the energy of a gauche conformation is not too high in the case of a hydrocarbon chain. However, to obtain the correct density for a hydrocarbon liquid the van der Waals radius must be chosen larger. This conflict may be solved by a separate treatment of third-neighbor and other van der Waals interactions.
4. There is some conflict between the wish to stick to conceptual simplicity of a force field on the one hand, and the wish to extend its range of applicability and accuracy by allowing the force field to become more complex.
5. Approximations inherent in certain force fields may block substantial improvement by changing of the parameters. For example, when applying a cut-off radius to long-range Coulomb forces, it is impossible to obtain a proper representation of electrostatic effects.

The inclusion of polarizability will be essential for an accurate modeling of processes like the binding of charged ligands. Polarizability only allows for a local displacement of charge, not for a transfer of charge. This can be achieved by combining static ab initio or semi-empirical quantum methods to compute charge densities with classical MD simulation in an adiabatic way: the quantum charge density is converted into atomic charges which are used in a number of classical simulation steps, and the new atomic positions are subsequently used to generate a new quantum charge distribution, and so on.<sup>[191]</sup> This type of treatment is only an improvement over conventional simulation when (i) the quantum Hamiltonian includes the long-range electrostatic

field, and (ii) the classical Hamiltonian accounts in a consistent way for the changing charge distribution.

In view of the limited accuracy of currently available force fields it may be wise in practical applications to check whether the results obtained are force field dependent.

#### 5.4. Non-Equilibrium Molecular Dynamics

In this paper we have considered the computer simulation of systems in equilibrium. It is also possible to change the equations of motion and boundary conditions such that the system is kept in a non-equilibrium state. In such a non-equilibrium molecular dynamics (NEMD) simulation, a non-equilibrium ensemble is sampled. NEMD simulation is an efficient technique to obtain transport coefficients, like viscosity, thermal conductivity, and mobility of molecular systems. An example is the study of the molecular viscosity of *n*-alkanes as a function of the intramolecular interaction function.<sup>[192]</sup> For an introduction to NEMD methods and their applications see Refs. [6, 7]. It is a rapidly developing field, which is still almost unexplored, with prospects in the area of practical rheology.

#### 5.5. Development of Computing Power

The growth of the field of computer simulation of fluid-like systems has been made possible by the steady and rapid increase of computing power over the last couple of decades. Figure 2 suggests an increase of an order of magnitude every 5–7 years. This trend will continue in the near future, since the present growth of computing power is based on the introduction of massive parallelism. The possibilities of parallel computation can be exploited rather easily in MD simulations, since the most time-consuming part is the force calculation, which can be carried out in parallel for all atoms in the system. Also the integration step can easily be performed in parallel. The required computing time for a MD simulation depends linearly on the simulation period and the number of particles (in case of forces of finite range). Therefore this type of calculation can optimally benefit from the continuous growth of computing power. In contrast, quantum mechanical calculations extend their range much more slowly, due to the fact that the required computing power is approximately proportional to  $N_e^3$  in the case of semi-empirical or local density functional calculations, to  $N_e^4$  in the case of Hartree-Fock calculations, or even to  $N_e^5$  in the case of configurational interaction calculations. Here,  $N_e$  is the number of *electronic* degrees of freedom. In classical simulations the required computing power is proportional to  $N_a$ , or in the case of inclusion of long-range forces or polarizability to at most  $N_a^2$ , where  $N_a$  denotes the number of *atomic* degrees of freedom.

### 6. Summary and Outlook

Dynamic computer simulation is just a branch of computational chemistry and physics in which a mathematical model of the real world is formulated and its consequences for the various physical or chemical quantities are evaluated

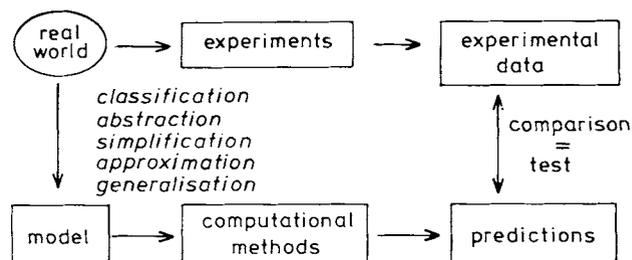


Fig. 17. Computational physics and chemistry involve formulation and testing of a (mathematical) *model* of the *real world*.

by numerical methods. This is illustrated in Figure 17. In this paper much attention has been devoted to methodology, because the core of any model is the approximations, assumptions and simplifications that are used in its formulation. Only an understanding of the basics of a particular model may lead to sensible application or improvement of its performance. Secondly, attention has been focussed on comparison of model predictions with experimental data, which may reveal flaws in the model (or experiment). Due to the complexity of the systems of chemical interest, theoretical methods only became of practical interest with the advent and development of computers. With the continuous progress of methodology and computing power, computational methods, especially computer simulation methods, will find wider and wider application in the different areas of chemistry.

In Table 7 we have listed our thoughts about the possible development of various aspects of (molecular dynamics)

Table 7. Development of various aspects of molecular dynamics computer simulation methods in chemistry.

| Aspect                      | Past (1980)   | Present (1990)  | Future (2000)                  |
|-----------------------------|---------------|-----------------|--------------------------------|
| accuracy                    |               |                 |                                |
| – atomic positions          | ≈ 3 Å         | ≈ 1 Å           | ≈ 0.5 Å                        |
| – free energy               | –             | ≈ 4 $k_B T$     | ≈ 2 $k_B T$                    |
| force field                 | united atoms  | all atoms       | polarizability                 |
| environment                 | vacuo         | solvents        | membranes                      |
| time scale                  | ≈ 10 ps       | ≈ 100 ps        | ≥ 1 ns                         |
| system size                 | ≈ 1000 atoms  | ≈ 10 000 atoms  | ≈ 100 000 atoms                |
| quantum (MD)                | –             | simple models   | enzyme reactions               |
| degrees of freedom          |               |                 |                                |
| non-equilibrium simulations | atomic fluids | simple polymers | rheology of molecular mixtures |

computer simulation methodology and application in chemistry. From the contents of this paper it will be clear that there are still a number of very difficult problems, such as protein folding or crystallization, which are well beyond the reach of simulation methods owing to the size of the configurational space involved, the time scale of the process, and the small free energy differences between folded and unfolded state or between crystalline and liquid state.

*We would like to thank all our co-workers and colleagues who contributed to the work described here. This work was*

supported by the Netherlands Foundation for Chemical Research (SON) with financial aid from the Netherlands Organisation for Scientific Research (NWO). The GROMOS (GRoningen MOlecular Simulation) program library, which can be used for applying the methods discussed here, is available at nominal cost from the authors.

Received: January 10, 1990 [A 776 IE]  
 German version: *Angew. Chem.* 102 (1990) 1020

- [1] R. W. Hockney, J. W. Eastwood: *Computer Simulation using Particles*. McGraw-Hill, New York 1981.
- [2] J. Hermans (Ed.): *Molecular Dynamics and Protein Structure*. Polycrystal Book Service, Western Springs, IL, USA 1985.
- [3] G. Ciccotti, W. G. Hoover (Eds.): *Molecular-Dynamics Simulation of Statistical-Mechanical Systems (Proc. Int. School Phys. "Enrico Fermi", Course 97)*, North-Holland, Amsterdam 1986.
- [4] D. L. Beveridge, W. L. Jorgensen (Eds.): *Computer Simulation of Chemical and Biomolecular Systems (Ann. N.Y. Acad. Sci. 482 (1986))*.
- [5] J. A. McCammon, S. C. Harvey: *Dynamics of Proteins and Nucleic Acids*, Cambridge University Press, London 1987.
- [6] M. P. Allen, D. J. Tildesley: *Computer Simulation of Liquids*. Clarendon, Oxford 1987.
- [7] G. Ciccotti, D. Frenkel, I. R. McDonald (Eds.): *Simulation of Liquids and Solids*, North-Holland, Amsterdam 1987.
- [8] W. F. van Gunsteren, P. K. Weiner (Eds.): *Computer Simulation of Biomolecular Systems*, Escom, Leiden 1989.
- [9] C. R. A. Catlow, S. C. Parker, M. P. Allen (Eds.): *Computer Modelling of Fluid Polymers and Solids (NATO ASI Ser. C 293 (1990))*.
- [10] M.-L. Saboungi, A. Rahman, J. W. Halley, M. Blander, *J. Chem. Phys.* 88 (1988) 5818.
- [11] W. L. Jorgensen, J. Tirado-Rives, *J. Am. Chem. Soc.* 110 (1988) 1657.
- [12] W. F. van Gunsteren, H. J. C. Berendsen: *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos, Groningen 1987.
- [13] N. L. Allinger, *J. Am. Chem. Soc.* 99 (1977) 8127; U. Burkert, N. L. Allinger: *Molecular Mechanics*, American Chemical Society, Washington D.C. 1982.
- [14] M. Levitt, A. Warshel, *Nature (London)* 253 (1975) 694.
- [15] B. J. Berne in: J. U. Brackbill, B. I. Cohen (Eds.): *Multiple Time Scales*, Academic, New York 1985, p. 419.
- [16] J. A. McCammon, M. Karplus, *Proc. Natl. Acad. Sci. USA* 76 (1979) 3585; *Biopolymers* 19 (1980) 1375.
- [17] B. J. Alder, T. E. Wainwright, *J. Chem. Phys.* 27 (1957) 1208; *ibid.* 31 (1959) 459.
- [18] A. Rahman, *Phys. Rev.* 136 (1964) A 405.
- [19] A. Rahman, F. H. Stillinger, *J. Chem. Phys.* 55 (1971) 3336.
- [20] L. V. Woodcock, *Chem. Phys. Lett.* 10 (1971) 257.
- [21] J. P. Ryckaert, A. Bellemans, *Chem. Phys. Lett.* 30 (1975) 123.
- [22] J. A. McCammon, B. R. Gelin, M. Karplus, *Nature (London)* 267 (1977) 585.
- [23] P. van der Ploeg, H. J. C. Berendsen, *J. Chem. Phys.* 76 (1982) 3271.
- [24] W. F. van Gunsteren, H. J. C. Berendsen, J. Hermans, W. G. J. Hol, J. P. M. Postma, *Proc. Natl. Acad. Sci. USA* 80 (1983) 4315.
- [25] W. F. van Gunsteren, H. J. C. Berendsen, R. G. Geurtsen, H. R. J. Zwinderman, *Ann. N.Y. Acad. Sci.* 482 (1986) 287.
- [26] J. de Vlieg, H. J. C. Berendsen, W. F. van Gunsteren, *Proteins* 6 (1989) 104.
- [27] G. Jacucci, A. Rahman, *J. Chem. Phys.* 69 (1978) 4117.
- [28] W. F. van Gunsteren in C. Troyanowsky (Ed.): *Modelling of Molecular Structures and Properties*, Elsevier, Amsterdam 1990.
- [29] J. E. H. Koehler, W. Saenger, W. F. van Gunsteren, *Eur. Biophys. J.* 15 (1987) 197; *ibid.* 15 (1987) 211.
- [30] J. Lautz, H. Kessler, R. Kaptein, W. F. van Gunsteren, *J. Comput. Aided Mol. Des.* 1 (1987) 219.
- [31] H. J. C. Berendsen, W. F. van Gunsteren, H. R. J. Zwinderman, R. G. Geurtsen, *Ann. N.Y. Acad. Sci.* 482 (1986) 269.
- [32] A. P. Heiner, R. R. Bott, H. J. C. Berendsen, W. F. van Gunsteren, unpublished.
- [33] J. E. H. Koehler, W. Saenger, W. F. van Gunsteren, *J. Biomol. Struct. Dyn.* 6 (1988) 181.
- [34] J. E. H. Koehler, W. Saenger, W. F. van Gunsteren, *Eur. Biophys. J.* 16 (1988) 153.
- [35] R. R. Ernst, G. Bodenhausen, A. Wokaun: *Principles of Nuclear Magnetic Resonance in One and Two Dimensions*. Clarendon, Oxford 1987.
- [36] J. Seelig, *Q. Rev. Biophys.* 10 (1979) 353.
- [37] E. Egberts, H. J. C. Berendsen, *J. Chem. Phys.* 89 (1988) 3718.
- [38] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, J. Hermans in B. Pullman (Ed.): *Intermolecular Forces*, Reidel, Dordrecht 1981, p. 331.
- [39] H. J. C. Berendsen, J. R. Grigera, T. P. Straatsma, *J. Phys. Chem.* 91 (1987) 6269.
- [40] W. F. van Gunsteren, H. J. C. Berendsen in J. Hermans (Ed.): *Molecular Dynamics and Protein Structure*, Polycrystal Book Service, Western Springs, IL, USA 1985, p. 5.
- [41] T. P. Straatsma, H. J. C. Berendsen, *J. Chem. Phys.* 89 (1988) 5876.
- [42] P. Dauber-Osguthorpe, V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest, A. T. Hagler, *Proteins* 4 (1988) 31.
- [43] F. A. Mommany, R. F. McGuire, A. W. Burgess, H. A. Scheraga, *J. Phys. Chem.* 79 (1975) 2361.
- [44] M. Levitt, *J. Mol. Biol.* 168 (1983) 595.
- [45] Z. I. Hodes, G. Nemethy, H. A. Scheraga, *Biopolymers* 18 (1979) 1565.
- [46] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, M. Karplus, *J. Comput. Chem.* 4 (1983) 187.
- [47] S. J. Weiner, P. A. Kollman, D. T. Nguyen, D. A. Case, *J. Comput. Chem.* 7 (1986) 230.
- [48] A. T. Hagler, E. Huler, S. Lifson, *J. Am. Chem. Soc.* 96 (1974) 5319.
- [49] J. Hermans, H. J. C. Berendsen, W. F. van Gunsteren, J. P. M. Postma, *Biopolymers* 23 (1984) 1513.
- [50] E. Platt, B. Robson, *J. Theor. Biol.* 96 (1982) 381.
- [51] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.* 79 (1983) 962.
- [52] D. W. Rebertus, B. J. Berne, D. Chandler, *J. Chem. Phys.* 70 (1979) 3395.
- [53] S. Lifson, A. T. Hagler, P. Dauber, *J. Am. Chem. Soc.* 101 (1979) 5111.
- [54] G. Nemethy, M. S. Pottle, H. A. Scheraga, *J. Phys. Chem.* 87 (1983) 1883.
- [55] L. G. Dunfield, A. W. Burgess, H. A. Scheraga, *J. Phys. Chem.* 82 (1978) 2609.
- [56] K. Heinzinger in C. R. A. Catlow, S. C. Parker, M. P. Allen (Eds.): *Computer Modelling of Fluid Polymers and Solids (NATO ASI Ser. C 293 (1990))* p. 357.
- [57] A. Paskin, A. Rahman, *Phys. Rev. Lett.* 16 (1966) 300.
- [58] S. N. Ha, A. Giammona, M. Field, J. W. Brady, *Carbohydr. Res.* 180 (1988) 207.
- [59] D. van Belle, I. Couplet, M. Prevost, S. Wodak, *J. Mol. Biol.* 198 (1987) 721.
- [60] F. J. Vesely, *J. Comput. Phys.* 24 (1977) 361.
- [61] B. J. Alder, E. L. Pollock, *Annu. Rev. Phys. Chem.* 32 (1981) 311.
- [62] P. Barnes, J. L. Finney, J. P. Nicholas, J. E. Quinn, *Nature (London)* 282 (1979) 459.
- [63] A. Warshel, H. Levitt, *J. Mol. Biol.* 103 (1976) 227.
- [64] S. T. Russell, A. Warshel, *J. Mol. Biol.* 185 (1985) 389.
- [65] W. G. J. Hol, P. T. van Duynen, H. J. C. Berendsen, *Nature (London)* 273 (1978) 443.
- [66] H. J. C. Berendsen in J. Hermans (Ed.): *Molecular Dynamics and Protein Structure*. Polycrystal Book Service, Western Springs, IL, USA 1985, 18.
- [67] A. Warshel, S. T. Russell, *Q. Rev. Biophys.* 17 (1984) 283.
- [68] B. R. Brooks in K. F. Jensen, D. G. Truhlar (Eds.): *Supercomputer Research in Chemistry and Chemical Engineering (ACS Symp. Ser. 353 (1987) 123)*.
- [69] J. Aqvist, W. F. van Gunsteren, M. Leijonmarck, O. Tapia, *J. Mol. Biol.* 183 (1985) 461.
- [70] A. J. C. Ladd, *Mol. Phys.* 33 (1977) 1039; *ibid.* 36 (1978) 463.
- [71] H. L. Friedman, *Mol. Phys.* 29 (1975) 1533.
- [72] J. Warwicker, H. C. Watson, *J. Mol. Biol.* 157 (1982) 671.
- [73] J. P. van Eerden in W. Olson (Ed.): *Nucleic Acid Conformation and Dynamics (Rep. NATO/CECAM Workshop)*, CECAM, Orsay 1983, p. 61.
- [74] P. B. Shaw, *Phys. Rev. A* 32 (1985) 2476.
- [75] R. J. Zauhar, R. S. Morgan, *J. Mol. Biol.* 186 (1985) 815; *J. Comput. Chem.* 9 (1988) 171.
- [76] P. Ewald, *Ann. Phys. (Leipzig)* 64 (1921) 253.
- [77] S. W. de Leeuw, J. W. Perram, E. R. Smith, *Proc. R. Soc. London A* 373 (1980) 27.
- [78] D. M. Heyes, *J. Chem. Phys.* 74 (1981) 1924.
- [79] K. B. Wiberg, R. H. Boyd, *J. Am. Chem. Soc.* 94 (1972) 8426.
- [80] M. Lipton, W. C. Still, *J. Comput. Chem.* 9 (1988) 343.
- [81] J. Moulton, M. N. G. James, *Proteins* 1 (1986) 146.
- [82] K. D. Gibson, H. A. Scheraga, *J. Comput. Chem.* 8 (1987) 826.
- [83] S. H. Northrup, J. A. McCammon, *Biopolymers* 19 (1980) 1001.
- [84] G. M. Crippen in D. Bawden (Ed.): *Chemometric Research Studies Series*, Wiley, New York 1983.
- [85] T. Havel, I. D. Kuntz, G. M. Crippen, *Bull. Math. Biol.* 45 (1983) 665.
- [86] T. Havel, K. Wüthrich, *Bull. Math. Biol.* 46 (1984) 673.
- [87] R. P. Sheridan, R. Nilakantan, J. S. Dixon, R. Venkataraghavan, *J. Med. Chem.* 29 (1986) 899.
- [88] A. DiNola, H. J. C. Berendsen, O. Edholm, *Macromolecules* 17 (1984) 2044.
- [89] M. Berkowitz, J. A. McCammon, *Chem. Phys. Lett.* 90 (1982) 215.
- [90] W. F. van Gunsteren, H. J. C. Berendsen, J. A. C. Rullmann, *Mol. Phys.* 44 (1981) 69.
- [91] Shi Yun-yu, Wang Lu, W. F. van Gunsteren, *Mol. Simulation* 1 (1988) 369.
- [92] T. Akesson, B. Jönsson, *Mol. Phys.* 54 (1985) 369.

- [93] S. Kirkpatrick, C. D. Gelatt Jr., M. P. Vecchi, *Science (Washington D.C.)* 220 (1983) 671.
- [94] R. A. Donnelly, J. W. Rogers, Jr., *Int. J. Quantum Chem.: Quantum Chem. Symp.* 22 (1988) 507.
- [95] Z. Li, H. A. Scheraga, *Proc. Natl. Acad. Sci. USA* 84 (1987) 6611.
- [96] W. Braun, N. Go, *J. Mol. Biol.* 186 (1985) 611.
- [97] W. F. van Gunsteren, M. Karplus, *Biochemistry* 21 (1982) 2259.
- [98] G. L. Seibel, U. C. Singh, P. A. Kollman, *Proc. Natl. Acad. Sci. USA* 82 (1985) 6537.
- [99] K. Remerie, W. F. van Gunsteren, J. B. F. N. Engberts, *Chem. Phys.* 101 (1986) 27.
- [100] M. J. Mandell, *J. Stat. Phys.* 15 (1976) 299.
- [101] L. R. Pratt, S. W. Haan, *J. Chem. Phys.* 74 (1981) 1864, 1873.
- [102] S. Gupta, J. M. Haile, *Chem. Phys.* 83 (1984) 481.
- [103] D. J. Adams, *Chem. Phys. Lett.* 62 (1979) 329.
- [104] J. B. Gibson, A. N. Goland, M. Milgram, G. H. Vineyard, *Phys. Rev.* 120 (1960) 1229.
- [105] C. L. Brooks III, M. Karplus, *J. Chem. Phys.* 79 (1983) 6312.
- [106] A. T. Brünger, C. L. Brooks III, M. Karplus, *Chem. Phys. Lett.* 105 (1984) 495.
- [107] C. L. Brooks III, A. Brünger, M. Karplus, *Biopolymers* 24 (1985) 843.
- [108] A. T. Brünger, C. L. Brooks III, M. Karplus, *Proc. Natl. Acad. Sci. USA* 82 (1985) 8458.
- [109] W. G. Hoover, D. J. Evans, R. B. Hickman, A. J. C. Ladd, W. T. Ashurst, B. Moran, *Phys. Rev. A* 22 (1980) 1690.
- [110] D. J. Evans, *J. Chem. Phys.* 78 (1983) 3297.
- [111] J. M. Haile, S. Gupta, *J. Chem. Phys.* 79 (1983) 3067.
- [112] S. Nosé, *Mol. Phys.* 52 (1984) 255.
- [113] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, J. R. Haak, *J. Chem. Phys.* 81 (1984) 3684.
- [114] H. C. Andersen, *J. Chem. Phys.* 72 (1980) 2384.
- [115] D. M. Heyes, *Chem. Phys.* 82 (1983) 285.
- [116] T. Schneider, E. Stoll, *Phys. Rev. B* 17 (1978) 1302; *ibid.* B 18 (1978) 6468.
- [117] D. J. Evans, G. P. Morriss, *Chem. Phys.* 77 (1983) 63; *Comput. Phys. Rep.* 1 (1984) 297.
- [118] J. M. Haile, H. W. Graben, *J. Chem. Phys.* 73 (1980) 2421.
- [119] M. Parrinello, A. Rahman, *Phys. Rev. Lett.* 45 (1980) 1196; *J. Appl. Phys.* 52 (1981) 7182; *J. Chem. Phys.* 76 (1982) 2662.
- [120] M. Parrinello, A. Rahman, P. Vashishta, *Phys. Rev. Lett.* 50 (1983) 1073.
- [121] S. Nosé, M. L. Klein, *Mol. Phys.* 50 (1983) 1055.
- [122] W. H. Press, B. P. Flannery, S. A. Teukolsky, W. T. Vetterling: *Numerical Recipes*, Cambridge University Press, Cambridge, England 1986.
- [123] C. W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice-Hall, Englewood Cliffs, NJ, USA 1971.
- [124] L. Verlet, *Phys. Rev.* 159 (1967) 98.
- [125] D. Beeman, *J. Comput. Phys.* 20 (1976) 130.
- [126] W. F. van Gunsteren, H. J. C. Berendsen, *Mol. Phys.* 34 (1977) 1311.
- [127] H. J. C. Berendsen, W. F. van Gunsteren in G. Ciccotti, W. G. Hoover (Eds.): *Molecular-Dynamics Simulation of Statistical-Mechanics Systems*, North-Holland, Amsterdam 1986, p. 43.
- [128] W. F. van Gunsteren, H. J. C. Berendsen, *Mol. Simulation* 1 (1988) 173.
- [129] G. Ciccotti, M. Ferrario, J.-P. Rijckaert, *Mol. Phys.* 46 (1982) 875.
- [130] F. J. Vesely, *Mol. Phys.* 53 (1984) 505.
- [131] H. J. C. Berendsen, W. F. van Gunsteren in J. W. Perram (Ed.): *The Physics of Superionic Conductors and Electrode Materials (NATO ASI Ser. B92)* (1983) 221.
- [132] W. F. van Gunsteren, *Mol. Phys.* 40 (1980) 1015.
- [133] J. Barojas, D. Levesque, B. Quentrec, *Phys. Rev. A* 7 (1973) 1092.
- [134] P. S. Y. Cheung, *Chem. Phys. Lett.* 40 (1976) 19.
- [135] D. J. Evans, S. Murad, *Mol. Phys.* 34 (1977) 327.
- [136] M. Fixman, *J. Chem. Phys.* 69 (1978) 1527.
- [137] J. Wittenburg: *Dynamics of Systems of Rigid Bodies*, Teubner, Stuttgart 1977.
- [138] J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, *J. Comput. Phys.* 23 (1977) 327.
- [139] W. F. van Gunsteren, M. Karplus, *Macromolecules* 15 (1982) 1528.
- [140] G. Ciccotti, M. Ferrario, J. P. Ryckaert, *Mol. Phys.* 47 (1982) 1253.
- [141] W. F. van Gunsteren in W. F. van Gunsteren, P. K. Weiner (Eds.): *Computer Simulation of Biomolecular Systems*, Escom, Leiden 1989, p. 27.
- [142] W. B. Street, D. J. Tildesley, G. Saville, *Mol. Phys.* 35 (1978) 639.
- [143] O. Teleman, B. Jönsson, *J. Comput. Chem.* 7 (1986) 58.
- [144] W. F. van Gunsteren, H. J. C. Berendsen, F. Colonna, D. Perahia, J. P. Hollenberg, D. Lellouch, *J. Comput. Chem.* 5 (1984) 272.
- [145] R. Boelens, R. M. Scheek, J. H. van Boom, R. Kaptein, *J. Mol. Biol.* 193 (1987) 213.
- [146] B. W. Matthews, *Nature (London)* 335 (1988) 294.
- [147] M. H. Caruthers, *Acc. Chem. Res.* 13 (1980) 155.
- [148] J. H. Miller, *J. Mol. Biol.* 180 (1984) 205.
- [149] H. Schindler, J. Seelig, *J. Chem. Phys.* 59 (1973) 1841.
- [150] R. Kaptein, R. Boelens, R. M. Scheek, W. F. van Gunsteren, *Biochemistry* 27 (1988) 5389.
- [151] W. F. van Gunsteren, R. Kaptein, E. R. P. Zuiderweg in W. K. Olson (Ed.): *Nucleic Acid Conformation and Dynamics (Rep. NATO/CECAM Workshop)* CECAM, Orsay 1984, p. 70.
- [152] G. M. Clore, A. M. Gronenborn, *Protein Eng.* 1 (1987) 275.
- [153] J. Lutz, H. Kessler, J. M. Blaney, R. M. Scheek, W. F. van Gunsteren, *Int. J. Pept. Protein Res.* 33 (1989) 281.
- [154] M. Niiges, G. M. Clore, A. M. Gronenborn, *FEBS Lett.* 239 (1988) 29.
- [155] H. Kessler, C. Griesinger, J. Lutz, A. Müller, W. F. van Gunsteren, H. J. C. Berendsen, *J. Am. Chem. Soc.* 110 (1988) 3393.
- [156] A. E. Torda, R. M. Scheek, W. F. van Gunsteren, *Chem. Phys. Lett.* 157 (1989) 289.
- [157] A. T. Brünger, J. Kuriyan, M. Karplus, *Science (Washington D.C.)* 235 (1987) 458.
- [158] P. Gros, M. Fujinaga, A. Mattevi, F. M. D. Vellieux, W. F. van Gunsteren, W. G. J. Hol in J. Goodfellow, K. Henrick, R. Hubbard (Eds.): *Molecular Simulation and Protein Crystallography (Proc. Joint CCP4/CCP5 Study Weekend)* SERC, Daresbury 1989, p. 1.
- [159] P. Gros, M. Fujinaga, B. W. Dijkstra, K. H. Kalk, W. G. J. Hol, *Acta Crystallogr. B* 45 (1989) 488.
- [160] R. R. Bott, *World Biotech. Rep.* 2 (1985) 51.
- [161] W. F. van Gunsteren, H. J. C. Berendsen, *J. Mol. Biol.* 176 (1984) 559.
- [161] P. Krüger, W. Strassburger, A. Wollmer, W. F. van Gunsteren, *Eur. Biophys. J.* 13 (1985) 77.
- [163] J. E. H. Koehler, W. Saenger, W. F. van Gunsteren, *J. Mol. Biol.* 203 (1988) 241.
- [164] J. Lutz, H. Kessler, W. F. van Gunsteren, H. P. Weber, R. M. Wenger, *Biopolymers* (1990), in press.
- [165] D. L. Beveridge, F. M. DiCapua, *Ann. Rev. Biophys. Biophys. Chem.* 18 (1989) 431.
- [166] D. A. Matthews, J. T. Bolin, J. M. Burridge, D. J. Filman, K. W. Volz, J. Kraut, *J. Biol. Chem.* 260 (1985) 392.
- [167] D. A. Matthews, J. T. Bolin, J. M. Burridge, D. J. Filman, K. W. Volz, B. T. Kaufman, C. R. Beddell, J. N. Champness, D. K. Stammers, J. Kraut, *J. Biol. Chem.* 260 (1985) 381.
- [168] B. Klar, B. Hingerty, W. Saenger, *Acta Crystallogr. B* 36 (1980) 1154.
- [169] Z. Otwinowski, R. W. Schevitz, R. G. Zhang, C. L. Lawson, A. Joachimiak, R. Q. Marmorstein, B. F. Luisi, P. B. Sigler, *Nature (London)* 335 (1988) 321.
- [170] V. Vlachy, D. J. Haymet, *J. Chem. Phys.* 84 (1986) 5874.
- [171] J. Tallon, *Phys. Rev. Lett.* 57 (1986) 2427.
- [172] R. W. Impey, M. Sprik, M. L. Klein, *J. Chem. Phys.* 83 (1985) 3638.
- [173] F. J. Abraham, W. E. Rudge, D. J. Auerbach, S. W. Koch, *Phys. Rev. Lett.* 52 (1984) 445.
- [174] D. M. Heyes, M. Barber, J. H. R. Clarke, *Surf. Sci.* 105 (1981) 225.
- [175] V. Rossato, G. Ciccotti, V. Pontikis, *Phys. Rev. B* 33 (1986) 1860.
- [176] U. Landman, W. D. Luedtke, R. N. Barnett, C. L. Cleveland, M. W. Ribarsky, E. Arnold, S. Ramesh, H. Baumgart, A. Martinez, B. Khan, *Phys. Rev. Lett.* 56 (1986) 155.
- [177] J. G. Powles, G. Rickayzen, *Mol. Phys.* 38 (1979) 1875.
- [178] W. G. Gibson, *Mol. Phys.* 28 (1974) 793.
- [179] P. H. Berens, D. H. J. Mackay, G. M. White, K. R. Wilson, *J. Chem. Phys.* 79 (1983) 2375.
- [180] B. J. Berne, G. D. Harp, *Adv. Chem. Phys.* 17 (1970) 63.
- [181] J. A. Barker, *J. Chem. Phys.* 70 (1979) 2914.
- [182] D. Chandler, P. G. Wolynes, *J. Chem. Phys.* 74 (1981) 4078.
- [183] B. de Raedt, M. Sprik, M. Klein, *J. Chem. Phys.* 80 (1984) 5719.
- [184] C. Zheng, J. A. McCammon, P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* 86 (1989) 6441.
- [185] E. J. Heller, *J. Chem. Phys.* 62 (1975) 1544 and 64 (1976) 63.
- [186] N. Corbin, K. Singer, *Mol. Phys.* 46 (1982) 671.
- [187] K. Singer, W. Smith, *Mol. Phys.* 57 (1986) 761.
- [188] R. Carr, M. Parrinello, *Phys. Rev. Lett.* 55 (1985) 2471.
- [189] A. Selloni, P. Carnevali, R. Carr, M. Parrinello, *Phys. Rev. Lett.* 59 (1987) 823.
- [190] C. H. Bennett, *J. Comput. Phys.* 19 (1975) 297.
- [191] P. A. Bash, M. J. Field, M. Karplus, *J. Am. Chem. Soc.* 109 (1987) 8092.
- [192] J. H. R. Clarke, D. Brown, *J. Chem. Phys.* 86 (1987) 1542.